selected articles from the

GTE LENKURT

DEMODULATOR

Volume 3

selected articles from the

GTE LENKURT DEMODULATOR

Volume 3

GTE LENKURT
San Carlos, California

Printed in the United States of America

PREFACE

The GTE Lenkurt Demodulator is a bimonthly publication that deals with current developments in the telecommunications industry. **Demodulator** articles are tutorial in nature and are intended for a wide readership including technicians, engineers, and members of management. Periodically, some of the most popular articles that have appeared as past issues are assembled in one volume, complete with table of contents and index for easy reference; Volume 3 of **Selected Articles From the GTE Lenkurt Demodulator** is the third such collection.

This volume contains 51 articles published between the years 1971 and 1975, and includes such categories of subjects as multiplex technology, pulse code modulation, data transmission, general communications, and developments and design.

EDITOR

THE GTE LENKURT DEMODULATOR

CONTENTS

SECTION I

MULTIPLEX TECHNOLOGY

SECTION II

PULSE CODE MODULATION

SECTION III

MICROWAVE RADIO

SECTION VI

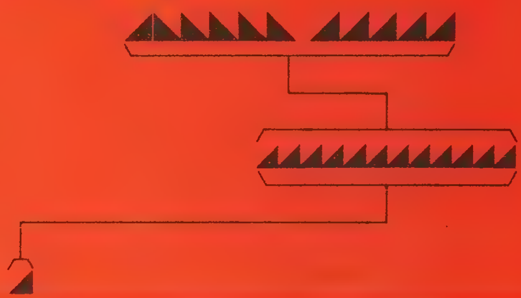
DEVELOPMENTS AND DESIGN

SECTION I
MULTIPLEX TECHNOLOGY

GTE LENKURT
DEMODULATOR
SEPTEMBER 1973

FDM Modulation Plans and Polylithic Crystal Filters

As developments in multiplex technology produce larger and more flexible systems, new modulation plans bring greater economic benefits to telecommunications equipment users.



Multiplexing is the means by which several circuits may be combined for transmission over a common transmission path. Modulation is a step in the process by which multiplexing is accomplished. While various types of modulation are possible, there are only two basic multiplexing techniques in common use – FDM (frequency division multiplexing) and TDM (time division multiplexing).

Frequency division multiplexing is a method of multiplexing in which two or more voice-frequency signals are translated to separate frequency bands by modulation processes so that they can then be combined and transmitted over a single medium. The two types of modulation used in FDM are amplitude modulation (AM) and frequency modulation (FM). This discussion centers around amplitude-modulated FDM systems.

Amplitude modulation is the process by which a carrier of constant frequency and constant amplitude is mixed with another signal, which is usually variable in both frequency and amplitude, and which is commonly called the modulating frequency. The amplitude of the carrier frequency is varied above and below its normal value in accordance with the modulating frequency (see Figure 1). The resultant of this process is three distinct and separate frequency components: (1) The original carrier-frequency component, (2) the sum of the carrier and the modulating frequency, and (3) the difference between the carrier and the modulating frequency. With a complex modulating wave, such as speech, the side frequencies above and below the carrier each consist of a band of frequencies. These two bands are called the upper and lower sidebands, and each contains the same information.

The sidebands obtained from a complex wave each have the same bandwidth as the original modulating wave; each contains the same intelligence; and, the frequencies in the upper sideband have the same relative relationship as the modulating wave; those in the lower sideband have an inverse relationship. Power distribution in the sidebands is directly related to the distribution of power in the modulating wave. As a result of amplitude modulation, the frequency band of the modulating wave is translated to a different position in the frequency spectrum. If both sidebands and the carrier are transmitted, the technique is called double-sideband transmitted-

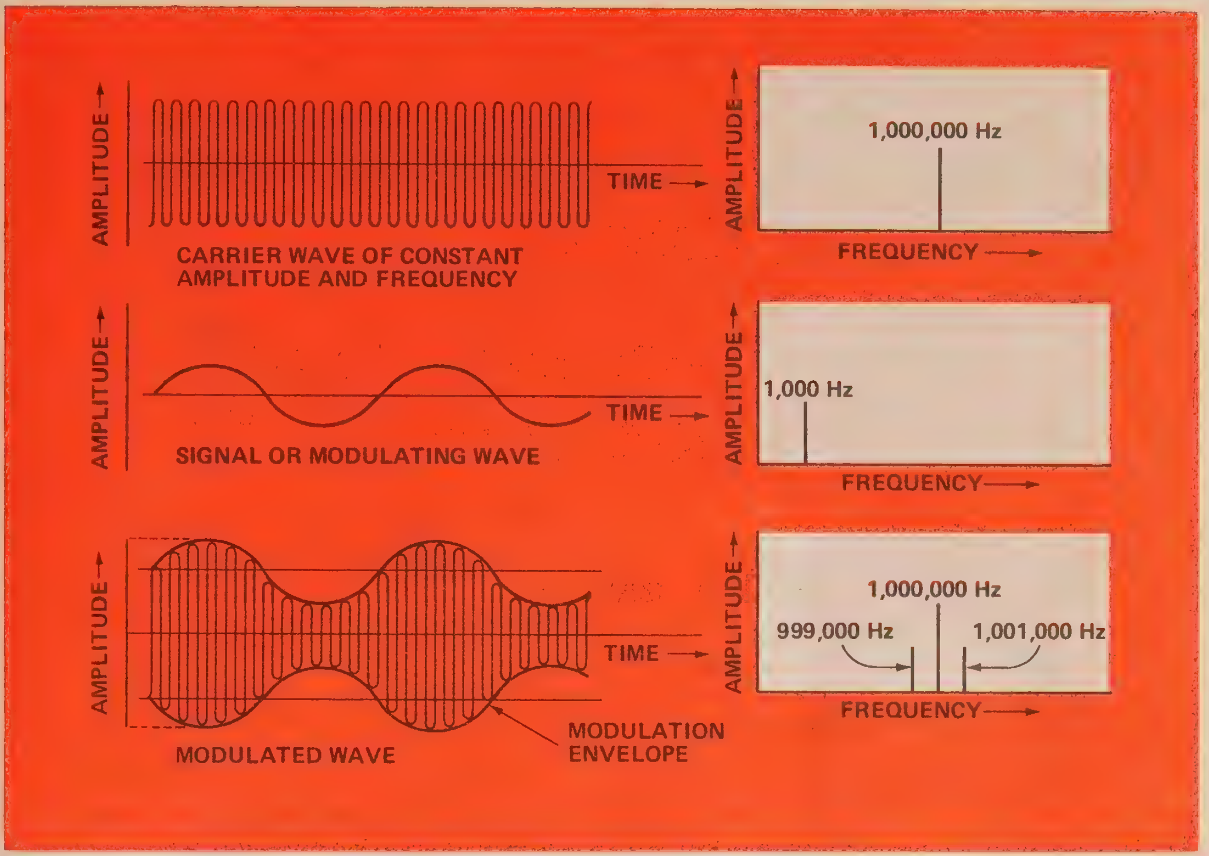


Figure 1. The resultant of amplitude modulation contains the carrier frequency component and the sum and difference of the carrier and modulating frequency.

carrier modulation (DSBTC). If the carrier is not transmitted, the technique is called double-sideband suppressed carrier (DSBSC).

Single-Sideband Suppressed Carrier

The most widely used modulation technique in frequency-division multiplexing is single-sideband suppressed carrier (SSBSC). Since the two sidebands produced during modulation contain identical information, only one is required to transmit the signal information. A balanced modulator is used to suppress the carrier, leaving only the two sidebands. Sideband suppression is accomplished by applying the modulated signal to a filter which passes only one of the sidebands while effectively attenuating the other. The bandwidth of the remaining sideband is now approximately equal to the original voice-frequency signal (about 3100 Hz).

Interconnection

When two carrier systems of different manufacture are interconnected within systems, signals at the interface point must satisfy the requirements of the receiving system. To this extent, it has been necessary to adopt standard modulation plans which allow different carrier and multiplex systems to be interconnected directly at line or baseband frequencies or at some intermediate stage of modulation. This allows groups of channels to be transferred between systems without the need for extra equipment and unnecessary modulation steps. Each type of carrier and multiplex system employs some type of modulation scheme to shift

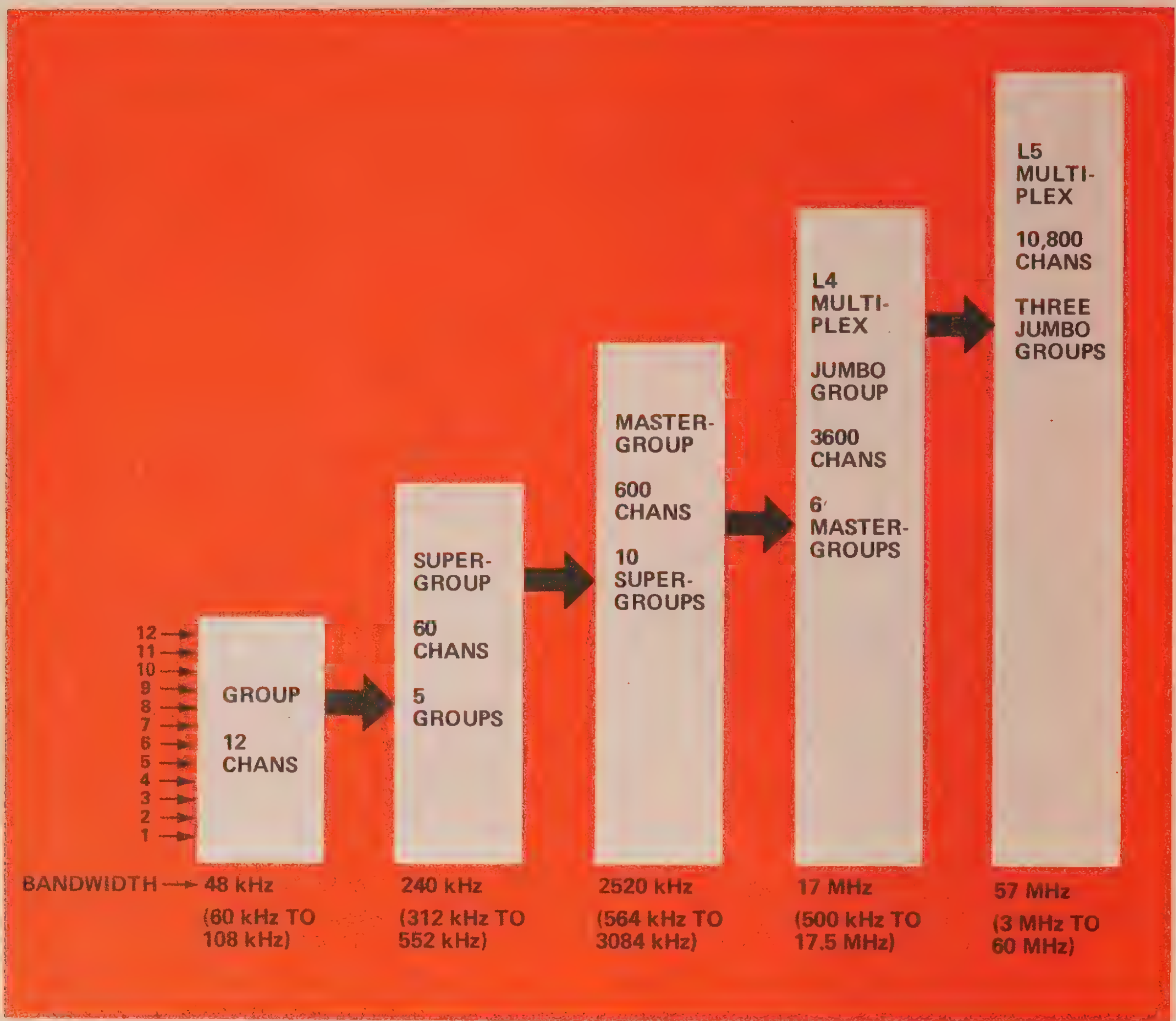


Figure 2. The hierarchy of FDM terminals.

the voice-frequency signals received from user equipment to a suitable line or baseband frequency range. This requires standardization of spacings of channel frequencies. In the 1930's the Bell System decided on a uniform channel spacing of 4 kHz; this requirement continues to be used today.

In SSBSC multiplex systems, voice channels are spaced at 4-kHz intervals to accommodate the voice frequencies in the range from about 300 Hz to 3400 Hz. Guardbands are provided between channels to prevent inter-channel crosstalk, and are also inserted between the various levels of the FDM hierarchy. To process a wider band of frequencies, such as required for video signals, multiplex systems often provide a means for combining groups of voice channels to acquire a single wideband channel.

FDM Terminals

Figure 2 shows the hierarchy of FDM terminals. The standard channel group contains 12 voice channels and extends in frequency from 60 kHz to 108 kHz (48-kHz bandwidth), with each channel occupying a band that is 4 kHz wide. The 12-channel group is widely accepted as the basic building block for long-haul carrier and multiplex systems. A standard 60-channel supergroup consists of five 12-channel groups, each at 60 kHz to 108 kHz.

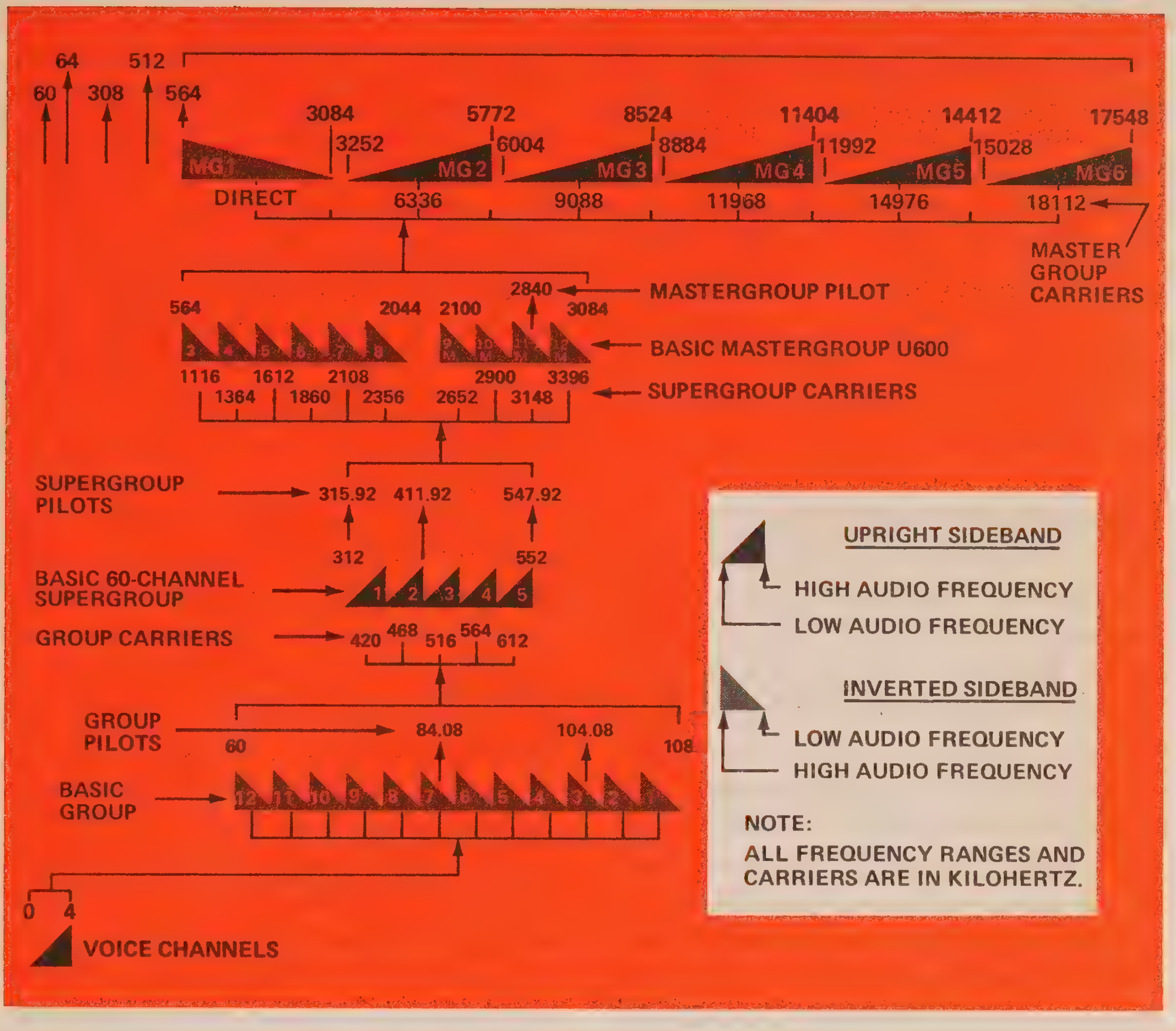


Figure 3. Conventional FDM frequency allocation and modulation.

The position of a 60 channel supergroup lies in the frequency spectrum of 312 kHz to 552 kHz. Ten supergroups make up a mastergroup which contains 600 channels, has a bandwidth of 2520 kHz, and lies in the 564 kHz to 3084 kHz region. The Bell System's L4 coaxial cable multiplex system assembles six mastergroups into a "jumbo" group to provide 3600 channels between 500 kHz and 17.5 MHz. And, in the future, three jumbo groups will be combined to provide 10,800 channels in the frequency spectrum between 3 MHz and 60 MHz.

An example of a conventional frequency allocation and modulation plan for a multiplex system is shown in Figure 3. In the first modulation stage, each voice frequency input signal modulates one of 12 channel carrier frequencies. The lower sideband signals are selected to provide the standard 60-kHz to 108-kHz, 12-channel group. In the second modulation stage, five 12-channel groups each modulate a separate group carrier frequency to produce a standard 60-channel supergroup with a frequency range of 312 kHz to 552 kHz. Ten of these supergroups form a 600-channel mastergroup. In the third stage of modulation, ten supergroups each modulate a separate supergroup carrier, resulting in line frequencies ranging from 564 kHz to 3084 kHz. Adherence to stan-

dard frequency allocation and modulation plans makes it possible to directly interconnect 12-channel, 60-channel, and higher-order channel groups of various carrier and multiplex systems, without having to first demodulate the signals down to the voice frequency range. With the large number of frequencies present in modern multiplex systems, frequency stability is a major concern to the user of telecommunications equipment.

Frequency Stability

The use of many individual carrier oscillators in a multiplex system would introduce a problem of frequency stability. Although frequency stability is often associated with the change in frequency of an oscillator over a period of time, the frequency stability of concern in a multiplex system is the net change in voice frequency that occurs between the sending end of the system and the receiving end. For voice communications via telephone systems, the permissable amount of frequency change and, indirectly, the required frequency stability, is related to the change which may occur without it being discernible to the ear. End-to-end frequency change of ±3 to ±5 Hz is satisfactory for voice communications. However, frequency stability is more critical in multiplex systems carrying telegraph and high-speed data. For this reason, most standards for frequency change in multiplex systems specify that the difference in frequency from one end of the system to the other shall not exceed 1 or 2 Hz.

Pilot Frequencies

Pilot frequencies are auxiliary signals employed in multiplex systems for such functions as level regulation, frequency synchronization, alarm systems, and maintenance monitoring. The transmit line levels at multiplex terminals and at repeaters must be maintained within close tolerances. Line noise and crosstalk from adjacent systems increase if the level is too low, while too high a level causes overloading, which can result in intermodulation distortion and crosstalk into other systems. Regulating pilots are used to operate compensating devices throughout the multiplex system in order to control line levels.

A line pilot is often used for end-to-end frequency synchronization. This is especially important in SSBSC systems, where the demodulating carrier frequencies must be reinserted at the receiving terminal. Frequency synchronization is accomplished by phase locking the master oscillator at a "slave" terminal to the line pilot frequency transmitted by the "master" terminal. Therefore, if the master oscillator frequency (thus, the line pilot frequency) at the transmit terminal changes, the synchronized oscillator frequency at the receive terminal will change a like amount.

The frequency of the pilots and the number required in each multiplex system depend mainly on the particular frequency allocation and modulation plan, and on any special needs of the system. Pilots are normally transmitted at a level 10 to 20 dB below the system test-tone level.

New Modulation Plans

Advances in filter technology have been a great influence in reducing the price and size of today's multiplex systems. Traditional LC filters have relatively low Q and have the disadvantage of limiting the passband due to their sloping sides. Although attenuation equalizers are used to "flatten" the passband of these filters, they add

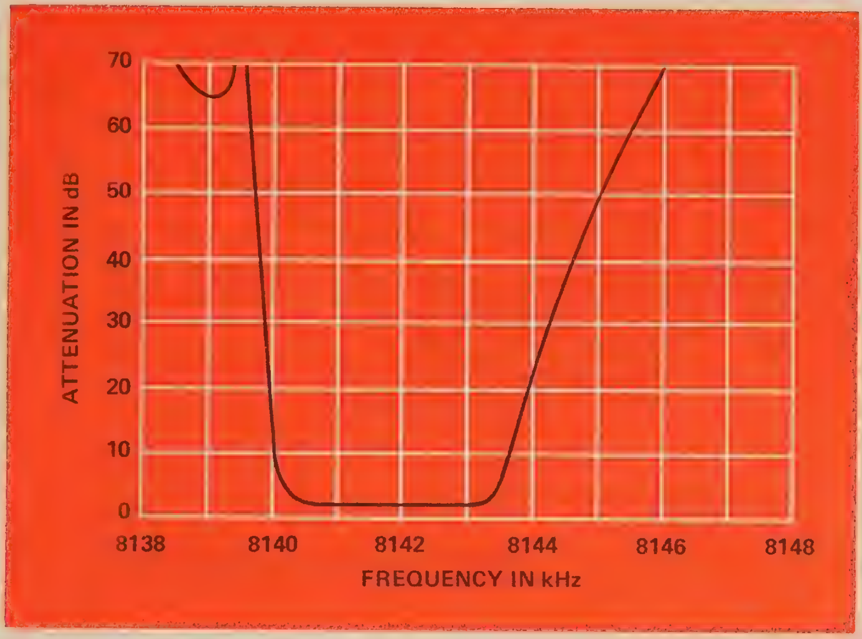


Figure 4. Typical passband of a polylithic crystal filter.

bulk and expense to the filter. The polylithic crystal filter, now being used in increasing proportions in more advanced systems, is a high Q device which provides a passband with steep sides and almost-square corners as shown in Figure 4. (See the November and December, 1970 issues of the Demodulator for a discussion of polylithic crystal filters.) The polylithic crystal filters manufactured by GTE Lenkurt operate most efficiently around 8 MHz, and it is within this frequency range that the latest modulation plans have been designed for new multiplex equipment such as the GTE Lenkurt 36A and 46A3. Both of these systems incorporate the crystal filter in their design although each system is intended for a different purpose. The 36A is a specialized direct-to-line system intended for the industrial user, while the 46A3 is a high-density toll-grade system designed for voice and data transmission over microwave radio or coaxial cable facilities. Both 36A and 46A3 systems may be used in combination, with the 36A occupying the lower frequency spectrum and the higher-density 46A3 occupying the spectrum above the 36A position.

The crystal filter used with the 46A3 allows the initial modulation and subsequent filtering to select the single sideband (upper) of the twelve 4-kHz channels to take place at 8140 kHz to 8188 kHz (48 kHz wide). The 8140-kHz to 8188-kHz primary group is then modulated with a single carrier, 8248 kHz for example, to place it in its standard group position (the 60-kHz to 108-kHz slot). Using this same technique, but different primary group modulating frequencies, DTL (direct-to-line) modulation can be accomplished.

Direct-To-Line Modulation

DTL modulation allows any 12-channel primary group to be positioned in any one of eleven desired 12-channel frequency allocations. This is done by mixing the 12-channel primary group (8140 kHz to 8188 kHz) with any of the eleven available carrier frequencies, as shown in the DTL portion of Figure 5. Use of the DTL technique allows the formation of a maximum of 132 channels that

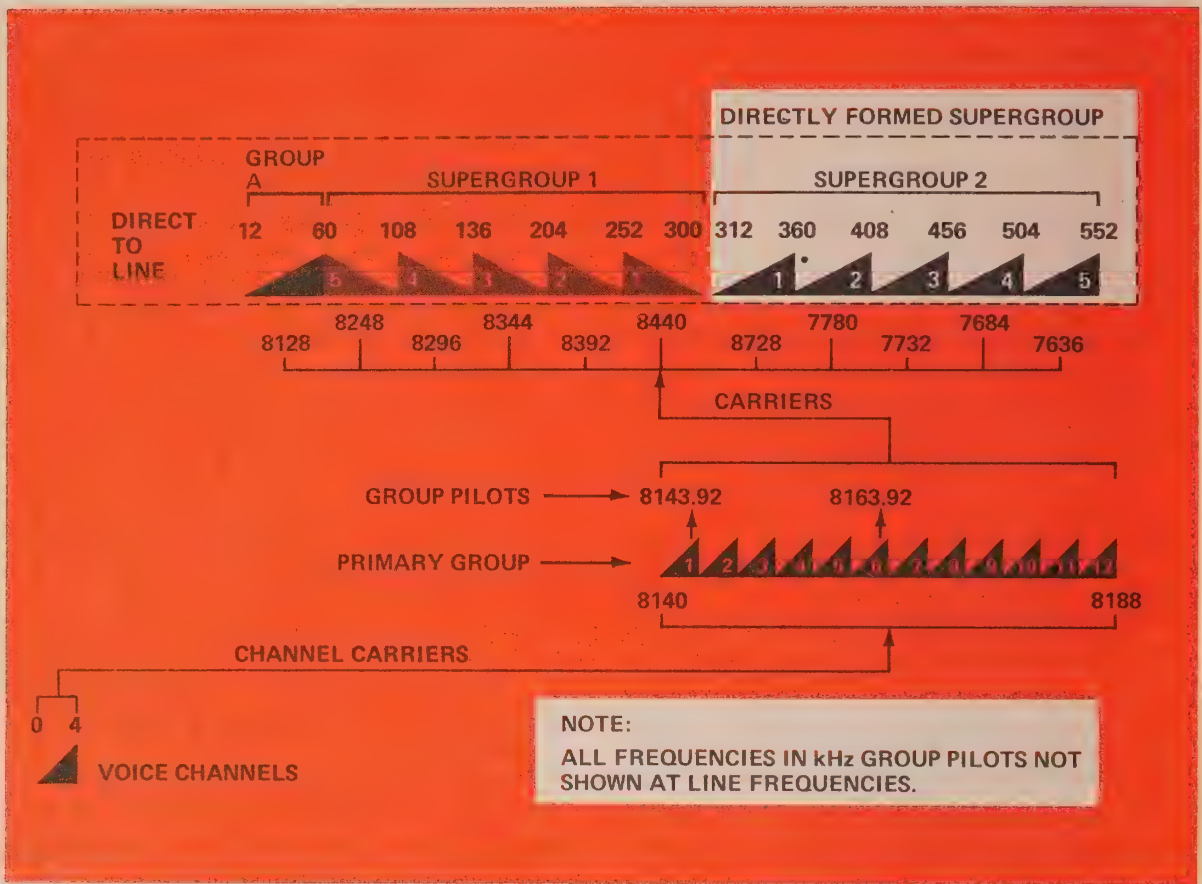


Figure 5. Direct-to-line and directly-formed supergroup modulation plans.

can be placed directly on a high frequency line in the band of 12 kHz to 552 kHz, without any other modulation required. This eliminates the group and supergroup equipment required in conventional modulation plans. DTL provides the correct sideband orientation to be compatible with conventional systems at the other end of a line and also eliminates one modulation step which results in better overall system performance. All frequencies are derived from a master-frequency source of 8192 kHz, which eliminates the possibility of frequency shifts that could cause channels to overlap and interfere with each other. The necessary carrier frequencies are obtained from the master source by binary chains of division and multiplication. DTL equipment offers the telephone, industrial, special service common carrier, and government users a low-cost, low-density, toll-grade carrier, which meets CCITT and U.S. standards. Also, the Group A portion of the DTL channel equipment conforms to the 12-kHz to 60-kHz channel group most often required by satellite communications systems. DTL equipment is designed for use in a microwave transmission system such as with the GTE Lenkurt 78F2 radio in the 2-GHz band. A block diagram of a typical DTL terminal is shown in Figure 6. For larger systems, the DFSG (directly formed supergroup) may be used.

Directly-Formed Supergroup

The DFSG forms a line spectrum of 312 kHz to 552 kHz, and can be used with either microwave radio or coaxial cable systems. DFSG equipment elimi-

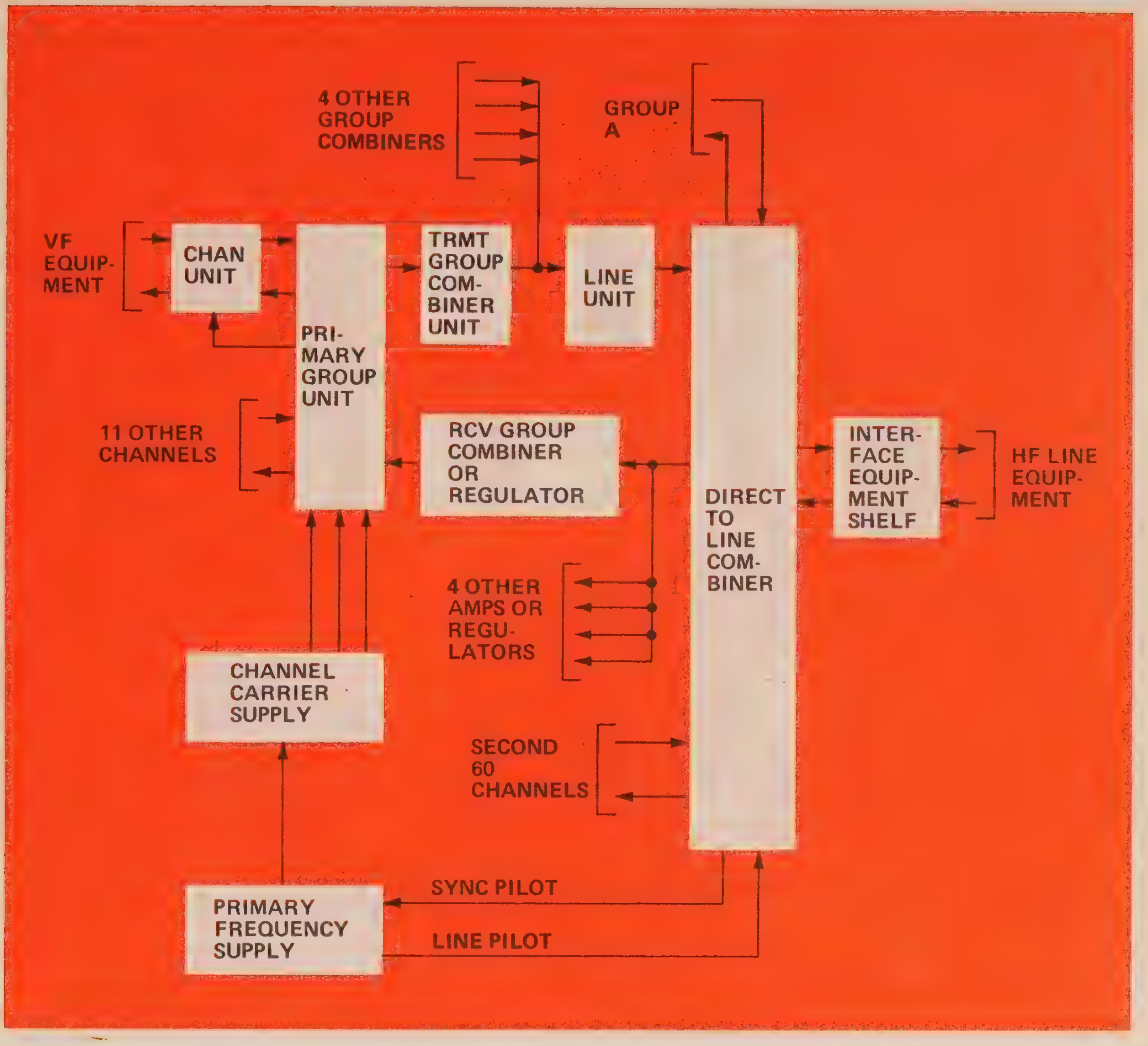


Figure 6. Typical DTL terminal configuration.

nates group shelves and group carrier supplies, and is designed for applications where a large number of channels is to be terminated at a single location. The DFSG can be fed to a basic supergroup shelf for a 600-channel system, which can in turn be translated to mastergroup equipment for a total of 1200 channels. Further developments will expand the 1200-channel capacity to 1800, 2400, and ultimately to a jumbo group of 3600 channels, and to combinations of jumbo groups. The modulation plan of the directly formed supergroup is shown in Figure 5. Twelve 0-4 kHz channels are arranged to form a primary group at 8140 kHz to 8188 kHz. Five primary groups are each modulated with a different carrier to form a basic 60-channel supergroup at 312 kHz to 552 kHz. This basic supergroup is formed directly with this one modulation step, in contrast to the conventional method of modulating five 60-kHz to 108-kHz basic groups with carrier frequencies to form a basic 60-channel supergroup.

The process of carrier frequency derivation requires that the 8192 kHz from the master frequency generator be mixed with a 4-kHz pulse, which will provide the 4-kHz frequency multiples necessary for the line fre-

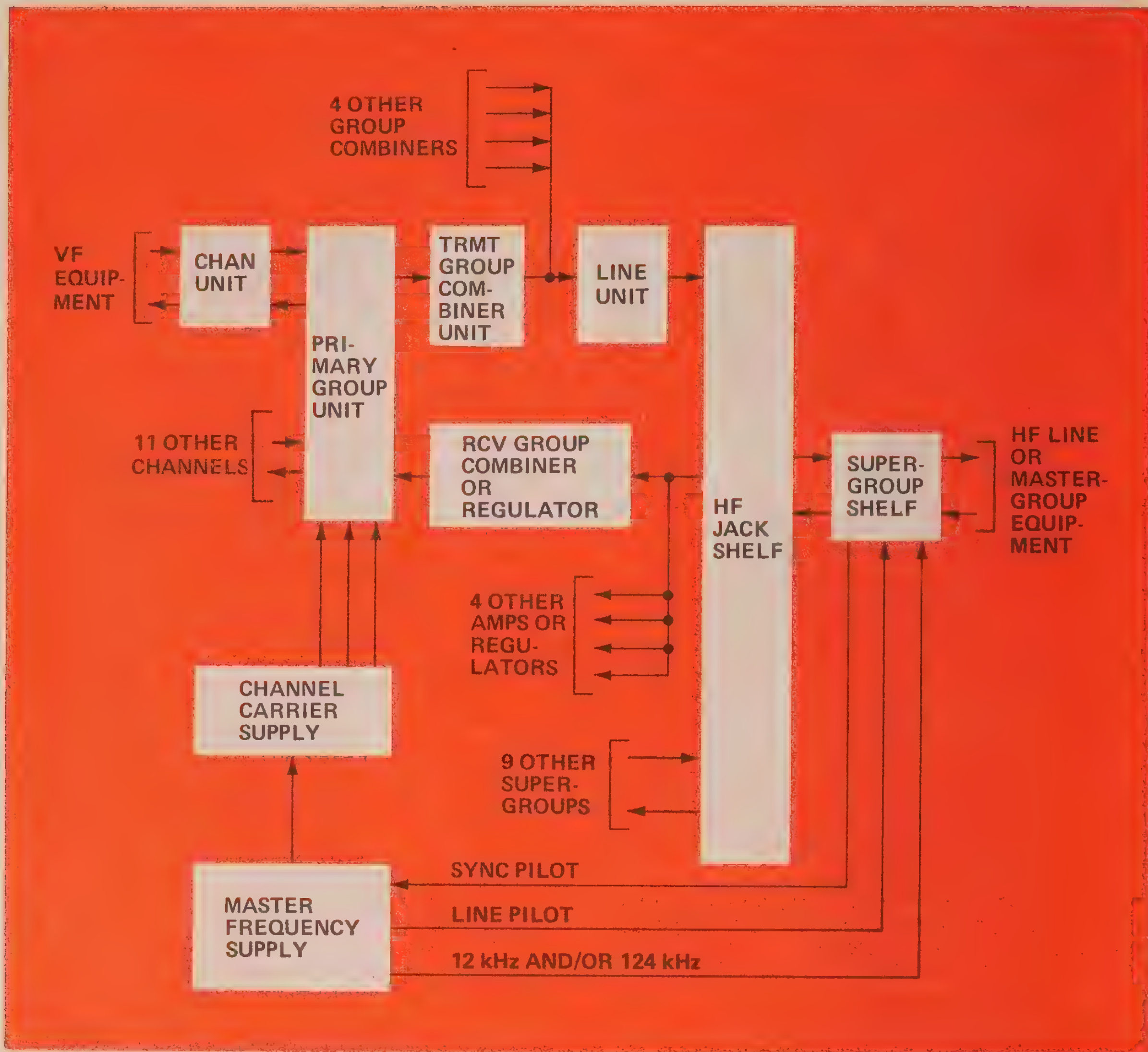


Figure 7. Typical DFSG terminal configuration.

quencies. The result of the mixing process produces a whole spectrum of frequencies at 4-kHz increments above and below 8192 kHz. This whole spectrum is fed into a narrow band crystal filter tuned to a particular frequency, and only that frequency will be passed. Here the polylithic crystal filter is used in a different application — as a narrow band pilot or carrier pick-off filter rather than a sideband bandpass filter.

Line pilots at any of five frequencies — 60 kHz, 64 kHz, 308 kHz, 512 kHz, or 564 kHz, may be used for frequency synchronization. The line pilot is used to phase lock the master frequency oscillator (8192 kHz), thereby preventing frequency shift from end-to-end. Figure 7 shows a block diagram of a typical DFSG terminal.

DTL and DFSG offer the advantage of flexibility for the high-density user. By proper system planning, a low-cost, initial system of DTL equipment may be installed and later expanded with DFSG or conventional channel equipment. Expansion into mastergroups is accomplished with conventional supergroup and mastergroup equipment. If 60 kHz to 108 kHz access is required within a given DFSG multiplex system, certain supergroups may be

equipped with conventional group equipment, thus mixing both DFSG and conventional equipments within the same system.

The elimination of group equipment through the use of DFSG equipment offers the user advantages over the use of conventional modulation steps. The first obvious advantage is a reduction in equipment costs – approximately 7% to 13%. Other advantages include, less equipment to install and maintain, increased reliability by reducing the amount of equipment required and improved noise performance for total systems.

DTL and DFSG offer new applications of GTE Lenkurt 46A3 and similar equipment. The direct-to-line equipment can meet the needs of the low-density systems, and directly-formed-supergroup equipment promises a savings in equipment and installation costs for high-density systems.

BIBLIOGRAPHY

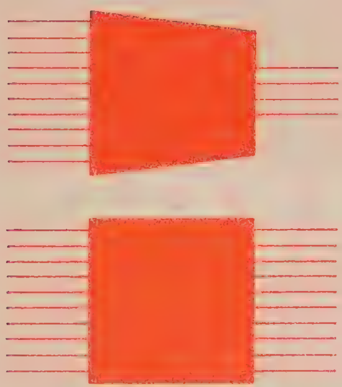
1. Kretzmer, Ernest R., "Communication Terminals," *Scientific American*, (September 1972), 131-140.

2. McLane, James A., "GTE Lenkurt 46A3 Direct-to-Line and Directly-Formed Supergroup Equipment," San Carlos, CA: GTE Lenkurt Inc., (July 1973).

3. Sheahan, Desmond F., "Single Sideband Filters For Short Haul Systems," *Proceedings of the IEEE*, (January 1971), 744.

GTE LENKURT
DEMODULATOR
AUGUST 1973

CONCENTRATORS & MULTIPLEXERS



The wires or lines that connect subscribers to the central office are the most inefficiently utilized portions of the telephone system. Concentrators and multiplexers are used to increase line efficiency.

Among the original patents on line concentrators, the following definition was given: "A line concentrator is essentially a switching device which provides for connections between a large plurality of subscriber lines and a small plurality of talking trunks." These subscriber lines can be either physical lines or carrier or multiplex channels. Similarly, a multiplexer is a device for combining a number of individual message circuits for transmission over a common transmission path. When used outside the central office (C.O.), both concentrators and multiplexers require fewer physical circuits to the central office than there are subscribers in the field.

For subscribers located at a considerable distance from the central office, the cost of physical circuits for each subscriber, and the time needed to install them, becomes quite substantial. This is especially true when dealing with a rapid influx of new subscribers. Both concentrators and multiplexers can provide this needed new service quickly and economically. But, they each provide this service in a different way and the grade of service is affected by different factors.

Multiplexers provide a trunk to the C.O. for each incoming voice or data channel by sharing frequency bands (FDM) or time slots (TDM) on a predetermined basis. Thus a multiplexer has the same instantaneous total input and output rate capacity. Concentrators, on the other hand, use a sharing or switching scheme in which some number of input channels share a smaller number of output channels on a demand basis. Consequently, it is not possible to have all concentrator subscribers using their phones simultaneously. For this reason, statistics and queuing play an important role in the planning and use of concentrators in an attempt to insure that trunks are available when needed.

Traffic Loading

Traffic loading has to be considered when planning to use a concentrator for service to remote locations. Telephone service is discussed in terms such as CCS, or erlang, for measuring traffic, and often refers to a grade of service as P(.01). The CCS stands for one hundred call-seconds, or the use of one telephone line for 100 seconds. Since there are 3,600 seconds in an hour, the maximum possible traffic rate on an individual line is 3,600 call-seconds per hour or 36 CCS (or 1 erlang) per hour – this would be continuous usage of the line. The term "P(.01) grade of service" refers to the probability that one call out of one hundred may be blocked due to lack of a free trunk. Similarly, a P(.02) would mean there is a probability of two calls out of one hundred being blocked; P(.03) of three; etc. Good

central office practices call for a P(.01) grade of service during the busy hour of the average day.

With a multiplexer, the grade of service of the C.O. is not affected and is only a function of the traffic loading. While with a concentrator, the grade of service is a function of the C.O. traffic loading *and* the traffic loading at the remote concentrator.

If a P(.01) grade of service is to be maintained, the number of subscribers that can be accommodated on a single concentrator system is limited, and is dependent upon the amount of telephone traffic that each subscriber generates. The actual traffic will vary from day to day and hour to hour, but there is an average that can be statistically determined. Service will only be degraded to more than one call in one hundred being blocked during the worst peaks of traffic, such as holidays and local emergencies when the calling rate exceeds this statistical average.

The number of subscribers that can be served by a concentrator is also dependent upon the type of service offered. For example, a concentrator system with a P(.01) grade of service designed to serve 96 single-party subscribers, under certain traffic conditions may only accommodate 72 two-party lines or 48 four-party lines. Combinations of single-, two-, and four-party lines can be served by the same concentrator, but the total number of lines offered decreases, for example, to 36 single-party lines, 24 two-party lines, and 24 four-party lines. Figure 1 shows the chart used to determine the number of subscribers in this specific example.

To minimize the number of cable pairs between a remote location and the central office, it is possible for a concentrator system to include a multiplexing technique — transmitting more than one signal on the same transmission channel. For example, a concentrator system that requires 24 cable pairs can serve the same number of customers with only 4 cable pairs if the concentrator is used in conjunction with a subscriber carrier system, such as the GTE Lenkurt 82A. Another concentrator arrangement uses a single (two pair) PCM (pulse code modulation) repeatered line. Figure 2 shows these concentrator arrangements.

A multiplex system such as the GTE Lenkurt 910A uses a PCM line to serve 24 customers over two cable pairs, or 48 customers over four cable pairs, and provides each of these customers with total access to the central office at all times. Each of these 24 single-party lines can be used for two- or four-party service, making it possible to increase the number of subscribers to 96 per two cable pair. Figure 3 shows the line arrangement using a multiplexer.

If the area served by a remote concentrator or multiplex system grows to the point where it is economical to put in a central office, an existing PCM multiplex system can be used to provide the trunks between central offices. The concentrator equipment may have to be removed and additional multiplex equipment put in for inter-office trunk requirements, if there are not enough cable pairs available.

Simultaneous Requests

Simultaneous originating and terminating requests made on a concentrator must be queued (lined up) and served preferentially. The concentrator may be simultaneously summoned by more than one line at the remote terminal requesting connection to the central office network. It is also possible for two or more callers to initiate termination requests that reach the

COMBINATIONS SHOWN IN THIS CHART ARE MAXIMUMS FOR P(.01) GRADE SERVICE WITH THE FOLLOWING BUSY-HOUR TRAFFIC:

ONE-PARTY UP TO 2.8 CCS
TWO-PARTY 3.41 - 4.0 CCS
FOUR-PARTY 5.41 - 6.5 CCS

NUMBER OF 2-PARTY LINES (rows); NUMBER OF 1-PARTY LINES (entries); NUMBER OF 4-PARTY LINES (columns)

NUMBER OF 2-PARTY LINES	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48
72	0																								
70	4	0																							
68	8	4	0																						
66	12	8	4	0																					
64	16	12	8	4	0																				
62	20	16	12	8	4	0																			
60	24	20	16	12	8	4	0																		
58	28	24	20	16	12	8	4	0																	
56	32	28	24	20	16	12	8	4	0																
54	36	32	28	24	20	16	12	8	4	0															
52	40	36	32	28	24	20	16	12	8	4	0														
50	44	40	36	32	28	24	20	16	12	8	4	0													
48	48	44	40	36	32	28	24	20	16	12	8	4	0												
46	50	47	43	39	35	31	27	23	19	15	11	7	3												
44	52	50	46	42	38	34	30	26	22	18	14	10	6	0											
42	54	52	49	45	41	37	33	29	25	21	17	13	9	3											
40	56	54	52	48	44	40	36	32	28	24	20	16	12	6	0										
38	58	56	54	51	47	43	39	35	31	27	23	19	15	9	3										
36	60	58	56	54	50	46	42	38	34	30	26	22	18	12	6	0									
34	62	60	58	56	53	49	45	41	37	33	29	25	21	15	9	3									
32	64	62	60	58	56	52	48	44	40	36	32	28	24	18	12	6	0								
30	66	64	62	60	58	55	51	47	43	39	35	31	27	21	15	9	3								
28	68	66	64	62	60	58	54	50	46	42	38	34	30	24	18	12	6	0							
26	70	68	66	64	62	60	57	53	49	45	41	37	33	27	21	15	9	3							
24	72	70	68	66	64	62	60	56	52	48	44	40	36	30	24	18	12	6	0						
22	74	72	70	68	66	64	62	59	55	51	47	43	39	33	27	21	15	9	3						
20	76	74	72	70	68	66	64	62	58	54	50	46	42	36	30	24	18	12	6	0					
18	78	76	74	72	70	68	66	64	61	57	53	49	45	39	33	27	21	15	9	3					
16	80	78	76	74	72	70	68	66	64	60	56	52	48	42	36	30	24	18	12	6	0				
14	82	80	78	76	74	72	70	68	66	63	59	55	51	45	39	33	27	21	15	9	3				
12	84	82	80	78	76	74	72	70	68	66	62	58	54	48	42	36	30	24	18	12	6	0			
10	86	84	82	80	78	76	74	72	70	68	65	61	57	51	45	39	33	27	21	15	9	3			
8	88	86	84	82	80	78	76	74	72	70	68	64	60	54	48	42	36	30	24	18	12	6	0		
6	90	88	86	84	82	80	78	76	74	72	70	67	63	57	51	45	39	33	27	21	15	9	3		
4	92	90	88	86	84	82	80	78	76	74	72	70	66	60	54	48	42	36	30	24	18	12	6	0	
2	94	92	90	88	86	84	82	80	78	76	74	72	69	63	57	51	45	39	33	27	21	15	9	3	
0	96	94	92	90	88	86	84	82	80	78	76	74	72	66	60	54	48	42	36	30	24	18	12	6	0

Figure 1. Traffic loading charts like this are used in determining the maximum number of subscribers that can be served by a concentrator.

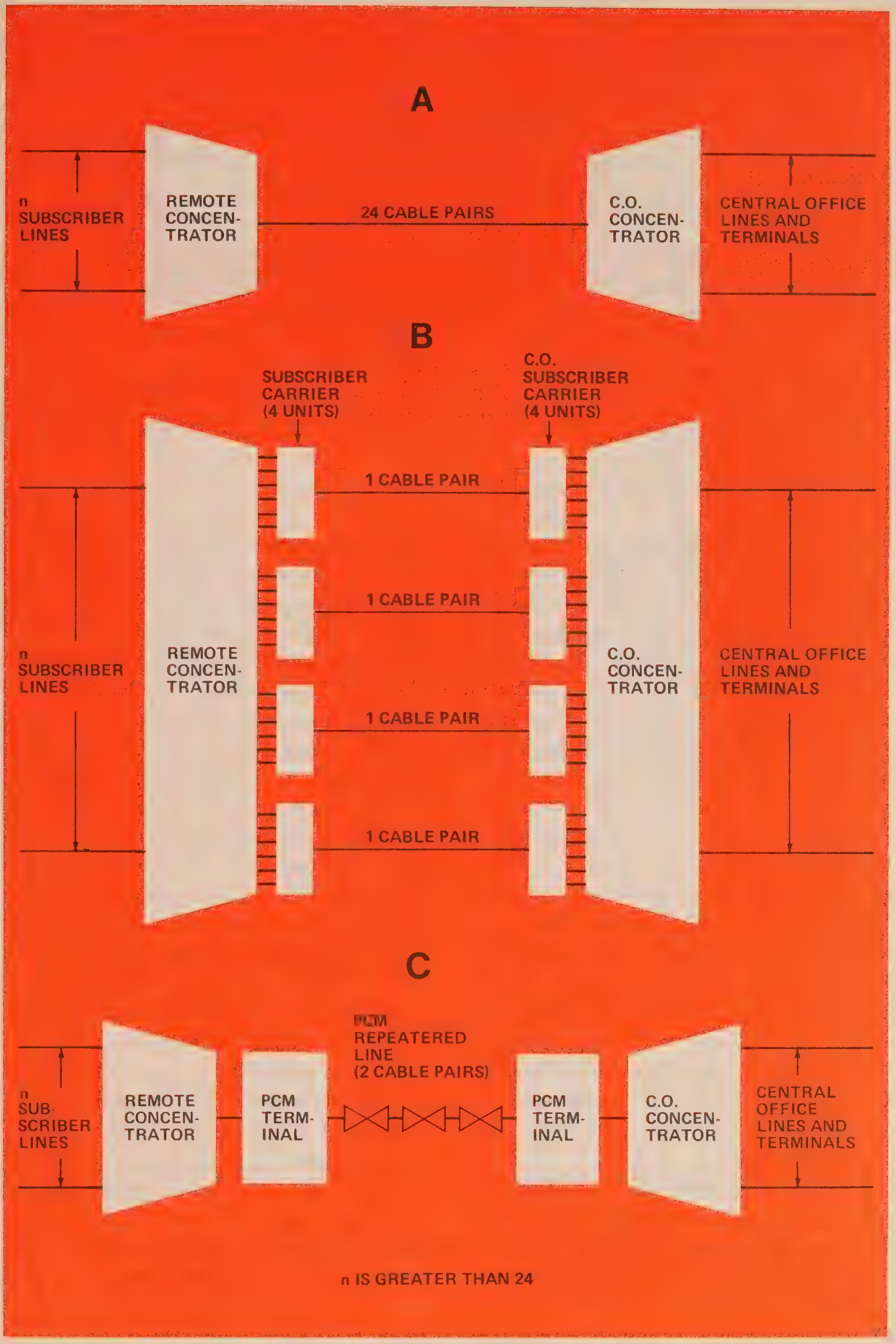


Figure 2. The number of cable pairs required for use with a concentrator varies with the addition of other equipment, such as a subscriber carrier system or a PCM repeatered line.

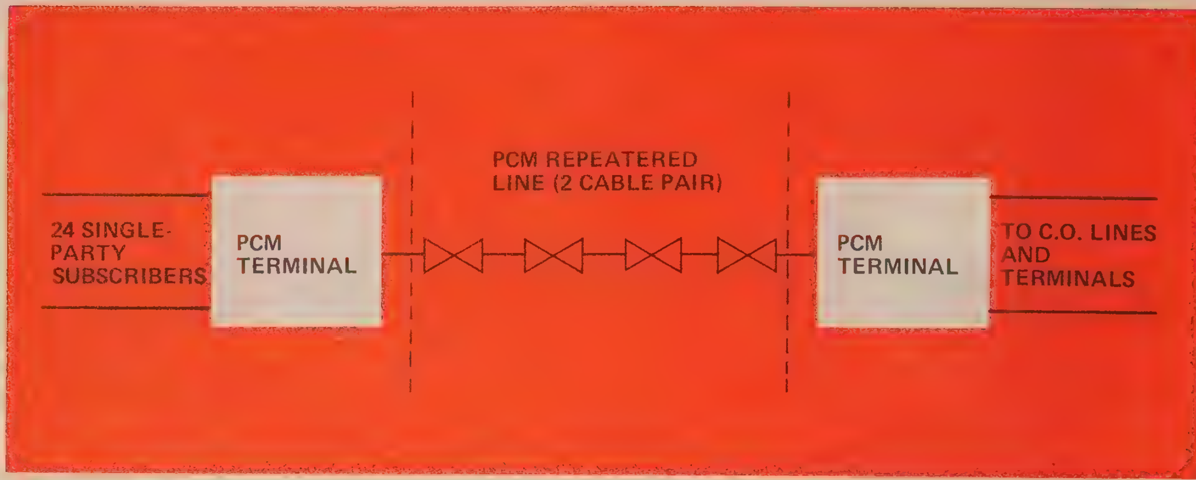


Figure 3. Using PCM terminals by themselves, it is possible to "concentrate" signals on two cable pairs but still provide total access to the central office by all subscribers.

concentrator simultaneously from the central office. If the concentrator is only able to serve one call at a time, a preference and lockout arrangement must be incorporated into the concentrator's design. Where the simultaneous requests originate at the same point, an electromechanical device is sufficient to provide the necessary preference and lockout. This switching device must be capable of recognizing the lines requesting service, determining which is to be served, remembering the identification of the served line, and preventing others waiting to be served from interfering. Concentrator switching is done in segments which simplifies the queuing process but limits the subscriber's access to the C.O. trunks. Instead of having access to all the C.O. trunks, he only has access to a fraction of them.

Some concentrators are also designed to recognize requests for service between two parties served by the same concentrator. Therefore it is not necessary to tie up central office equipment or the trunks to the central office, except for dialing. Once the C.O. terminal recognizes an intracall, the call is connected through at the remote terminal of the concentrator. Once these intratrunks are used up, each subsequent call between any other subscribers using the same concentrator must be connected through the C.O. switching equipment requiring two C.O. trunks. This could quickly degrade service if the area served by the remote concentrator functions as a small community, like a retirement community or a mobile home development where intracalls are frequent. Using a multiplex system, on the other hand, any two parties must be connected through the central office. This ties up the central office equipment, but would not prevent other subscribers at the remote terminals from reaching the central office, since each multiplex subscriber has his own trunk to the central office.

Assuming the called party's phone is not in use, the use of a multiplexer still does not guarantee that all calls will go through because concentration techniques are used throughout the switched telephone network, and have been for many years. It would be too costly and wasteful to have enough "lines" so that all telephone patrons could be talking simultaneously. Sta-

tistics have shown that even at peak traffic periods, telephone usage is only about 50%. It is partially because of this proven usefulness of concentration techniques within the central office that concentrators were proposed for use outside the central office. In remote locations where there is a sudden influx of subscribers, but where there is not enough growth to warrant a new central office, and where there is not enough capital or time to put in cable pairs for each subscriber, concentrators and multiplexers both seem to offer workable solutions.

Final Choice

The decision to purchase concentrators or multiplexers is based on a variety of factors regardless of whether the information being transmitted is voice or data. There are some applications where the choice is obvious because one or the other, but not both, will do the job required. With voice transmission, the choice usually depends on such subjective things as the quality of service – how often is a call actually blocked when using a concentrator? Statistics are fine for predicting the average traffic loading and the resulting quality of the service, but once installed, who's to say whether the actual use will be anything like the predicted average. With a multiplex system, each subscriber can have his own private line to the central office and the switched network, and statistical traffic loading at the subscriber end is no longer a consideration for the system planner. At this point, cost becomes a major factor, not only in the cost of the equipment but also in the facilities needed for installation, plans for future expansion, necessary routine maintenance, and periodic traffic studies. Both concentrators and multiplexers make more efficient use of cable plant, but in some applications, one will be more efficient and economical than the other.

BIBLIOGRAPHY

"Concentrator and Multiplexer: Discrete Terms for Discrete Functions." *Telecommunications.* January 1971.

Doll, Dixon R. "Multiplexing and Concentration." *Proceedings of the IEEE.* Vol. 60, no. 11. November 1972.

Joel, A. E. Jr. "An Experimental Remote Controlled Line Concentrator." *The Bell System Technical Journal.* March 1956.

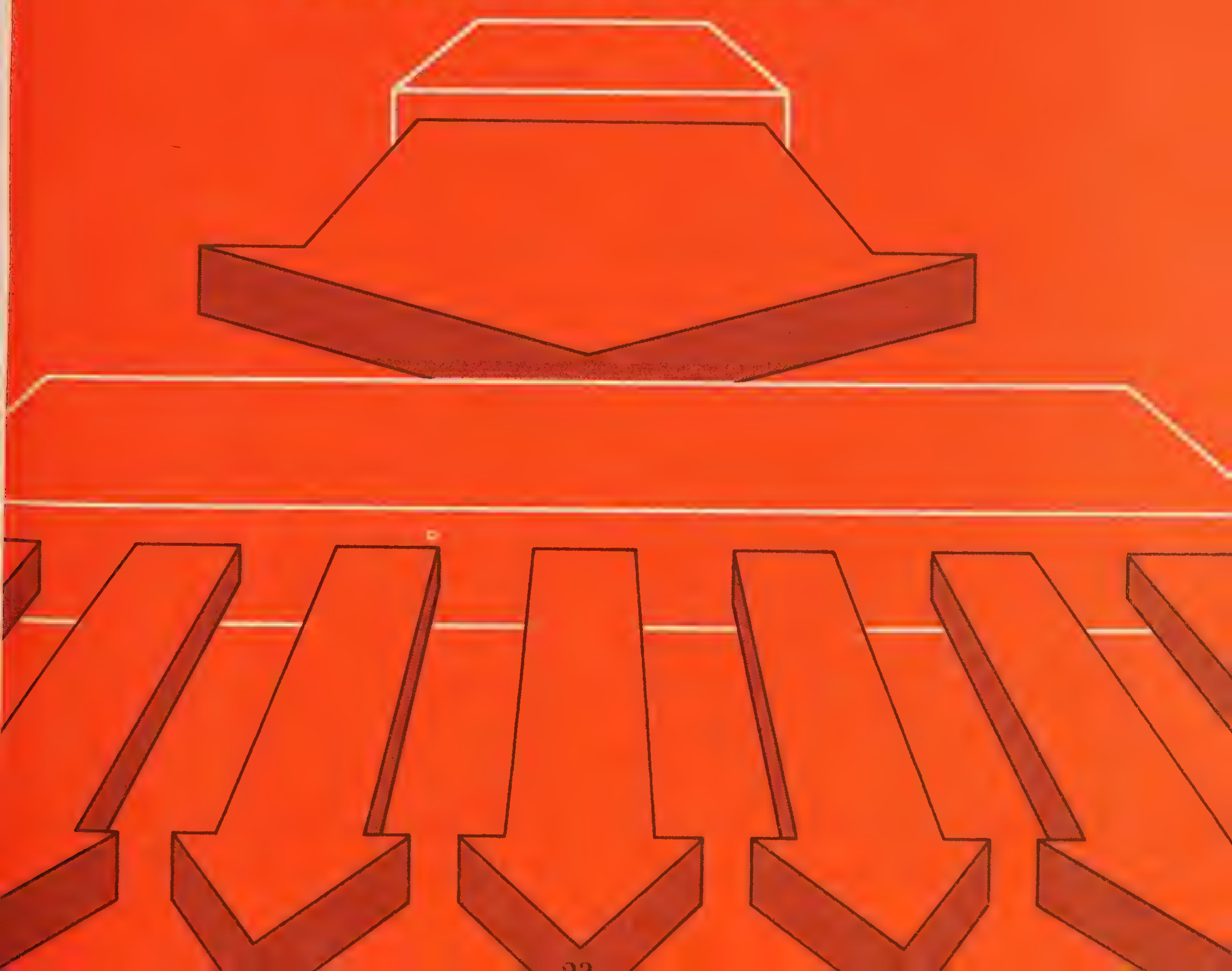
"Line Concentrators." Ericsson Review. No. 4 (1968).

Stagg, U. K. "The 2A Line Concentrator." *The Bell System Technical Journal.* October 1965.

U.S. Patent No. 2,806, 088. Granted 10 September 1957.

GTE LENKURT
DEMODULATOR
OCTOBER 1972

Carrier In Customer Distribution Networks



Once considered a temporary expedient for expanding the customer distribution network, subscriber carrier now offers permanent, improved service, and economic savings to both customer and telephone company.

Necessity gives impetus to invention, and it was necessity that led to the development of carrier systems specifically designed for the customer distribution network. Today, the result of this development is a potential savings of millions of dollars annually. The customer distribution network includes central office and switching equipment as well as all of the telephone equipment and lines from the central office to the customer's premises. It is also known as the exchange distribution network or the loop plant. Station or subscriber carrier equipment is located between the customer and the central office, in the area known as the customer or subscriber loop.

A carrier system on the customer's end allows two or more independent conversations to take place on one- or two-wire pairs that extend between telephone customers and the central office. Prior to the development of such carrier systems, each telephone line was linked to a central office by a single, physical wire pair, with power being supplied for the customer loop by the central office. (See Figure 1.)

The earliest form of what might be termed subscriber carrier, employed the technique of superimposing voice modulated carrier-frequency signals on electric power lines. This technique, called "power-line carrier," brought telephone service to remote customers (provided they had electricity) without the necessity of stringing telephone lines.

Equipment developed in the early 1950's was designed to assign several customers to each line, thereby providing new-customer service without adding cable pairs. Originally, this type of service, which postponed the high cost of installing new open wire, was primarily used to reach remote customers. These early systems provided service for 2 to 10 parties on one loop. The electronic circuitry of the early carriers made wide use of vacuum tubes and therefore required periodic maintenance by skilled personnel. Much of the maintenance of these systems consisted of replacing aging tubes or allowing for aging by adjustment of power levels on the line. For this reason, carrier has been mainly regarded as a temporary expedient for telephone service until physical lines can be installed. Today, with the era of the vacuum-tube carriers almost passé, solid-state subscriber carrier as a permanent medium, warrants investigation.

The many telephone companies already using permanent subscriber carrier have based their decision to do so on certain relevant facts. First, copper, being an element of the earth, has a finite quantity. As it becomes more scarce, its price will increase. Also, the labor cost of laying or stringing new cable can only go up in the future. Carrier may be used to circumvent some of these expenditures. The solid-state components of modern carrier equipment need little or no maintenance, and their performance and reliability far surpasses that of the early vacuum-tube equipment. Some additional points in favor of modern carrier systems as compared to older

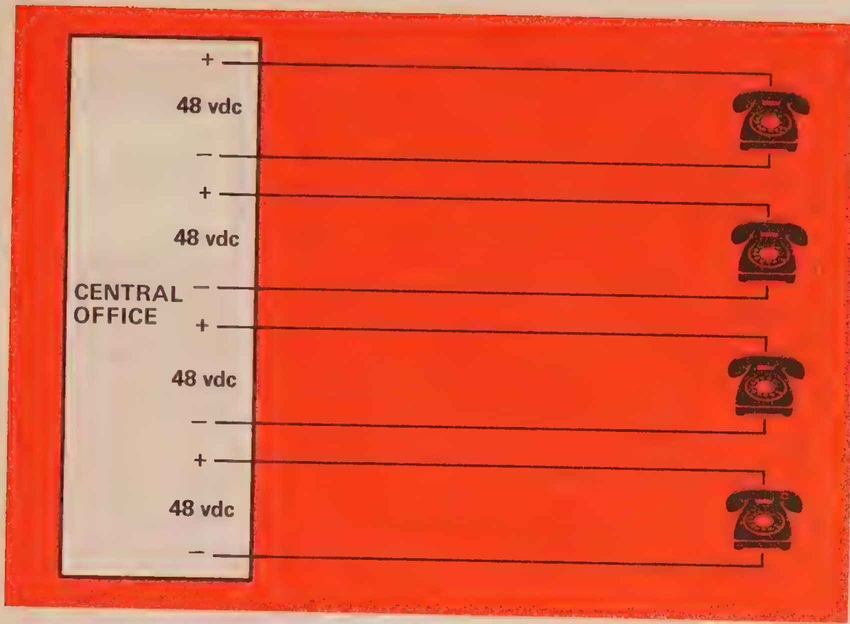


Figure 1. Prior to carrier, each subscriber line was powered by the central office and required one physical cable pair per customer circuit.

types include conservation of floor and rack space, lower terminal-equipment cost, lower power consumption, and lower repeater costs.

Subscriber and Station Carrier

The terms subscriber carrier and station carrier are often used synonymously because both systems perform the same function in much the same way. There is, however, by arbitrary definition, a technical difference between the two. In a subscriber carrier system, channels are usually terminated at one or two remote points, with the remote terminals being locally powered. Service to customers is then distributed from these remote points. On some systems, each channel drop can operate at loop resistances (total resistance of the length of wire between the subscriber's premises and the remote subscriber terminal) of up to 1600 ohms.

GTE Lenkurt recently developed a PCM (pulse code modulation) subscriber carrier system. PCM subscriber carrier systems such as the GTE Lenkurt 910A provide two groups of 24 single-party channels, and may be equipped with four-party channel units to serve 192 subscribers over four cable pairs.

Station carrier systems have been developed by several manufacturers, and although different design techniques have been used, the overall operation of the equipment is similar. The two basic choices of equipment are single-channel and multichannel carrier. Single-channel carrier provides an additional single-party line over one cable pair. Multichannel carrier provides up to 10 voice channels on one cable pair.

Station carrier systems permit individual remote channels to be located randomly anywhere along the carrier transmission path, since they are all powered from the serving central office. Typically, the channel drops (the length of line from the remote terminal to the customer) of these systems are limited to a few hundred feet of wire because of central office power limitations on the carrier equipment. Subscriber carrier systems generally require some initial adjustments by maintenance personnel, while station carrier systems automatically regulate both carrier and voice-frequency signals, and therefore require no additional adjustments. Figure 2 shows a typical example of subscriber and station carrier applications. For simplicity, this article will refer to station

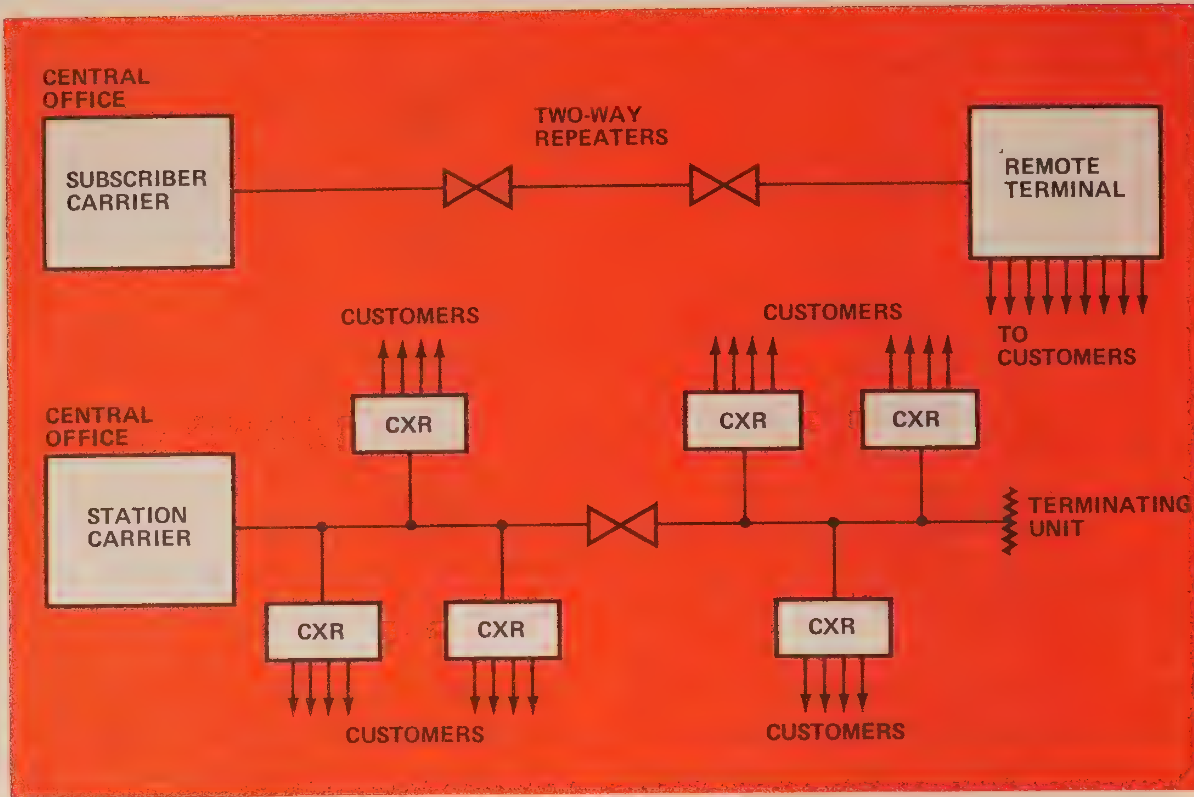


Figure 2. Subscriber carrier systems distribute customer lines from locally powered remote terminals. Station carrier terminals are powered from the central office and are randomly distributed along the line.

carrier and subscriber carrier, simply as *carrier*, unless a point must be made with reference to a particular system.

Uses

The same economic necessity that led to the initial development of subscriber carrier has increased in recent years with the migration of customers to suburban and rural areas. The multiparty line that was once considered adequate in rural areas is no longer acceptable to many new customers and the trend is to provide single or two-party service. This increasing demand for upgraded telephone service has depleted the stock of available cable pairs in many areas. Figure 3 shows a typical customer distribution network extending over a 15 to 20-mile distance with an almost-depleted cable supply at junction D. Should the area beyond junction D experience an influx of new customers, the decision of whether to install additional cable or use subscriber carrier must soon be made. Installation of cable is a costly venture in any case, and may often be forestalled or avoided by installing carrier equipment. Also, an error in estimation of future needs for that area can be costly. Too few cable pairs will necessitate additional cable installation in the future. Too many cable pairs lying idle over a period of years can mean economic waste in material and labor costs. Utilization of permanent carrier may reveal, in many cases, that additional cable will not be required for many years. In addition to eliminating the cost of cable installation, use of carrier avoids the time loss incurred between order and delivery of cable, and the scheduling of construction crews for new cable plant.

Carrier systems can also fill the requirement for short-notice demands for business, residential, or seasonal services. The simplicity of installation and maintenance is a great economic

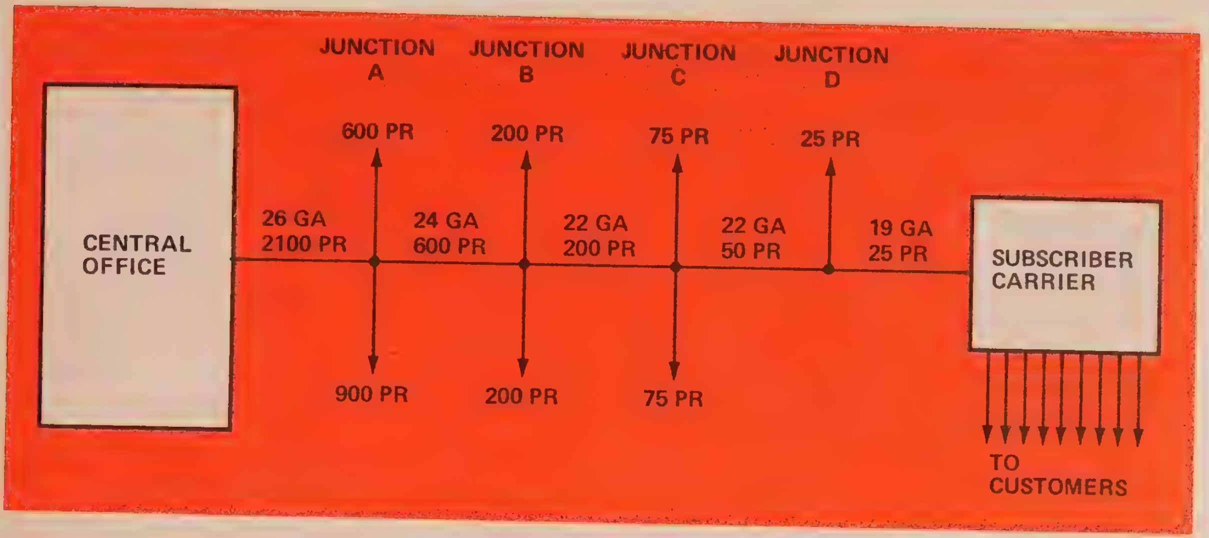


Figure 3. Addition of subscriber carrier to existing cable provides many new customer telephone channels.

benefit in such cases. As in much of today's electronic equipment, trouble-shooting of carrier systems consists of isolating a problem to a particular channel unit and replacing the inoperative unit with a spare.

Cable Condition

Carrier equipment can only be as good as the cable over which it transmits. The decision to install carrier on existing cable facilities must take into consideration the serviceability of the cable. Besides pair integrity, cable pairs to be used for carrier transmission must be devoid of loading coils, building-out networks, and bridge taps. (See the December, 1971 issue of the Demodulator for one method of detecting these devices.) On multi-channel station carrier systems, it is also necessary to install a terminating unit near the distant end of each cable pair to eliminate carrier frequency reflections which can cause signal attenuation. The maximum operating length of any carrier system depends on the gauge of wire that is used. On most multichannel station carrier systems, repeaters are required at approximately 35-dB intervals along the line.

The GTE Lenkurt 910A PCM system, which uses a D2-type format, can feed 15 repeaters from the central office and 15 more from the remote terminals for a total of 30 repeaters. (See the January, 1972 issue of the Demodulator for an explanation of D2-type terminals.) On 19 gauge cable, the typical spacing between these repeaters is about 8000 feet. This means that a fully equipped PCM system can extend to 240,000 feet (30 repeaters X 8000 ft.), or approximately 45 miles. The use of 24 gauge cable decreases repeater spacings to 5,000 feet, which gives a maximum range of approximately 28 miles. Greater distances, as much as 500 miles if necessary, can be achieved by addition of auxiliary power units along the line. As with other types of carrier systems, PCM channels can be added as they are needed, thus minimizing initial expenditures. PCM also provides an improved quality of service on long applications, since signals are regenerated at each repeater and retransmitted without any noise that may have been picked up along the line.

New Installation

The high reliability of modern subscriber and station carrier systems has made them acceptable as permanent installations rather than just tempor-

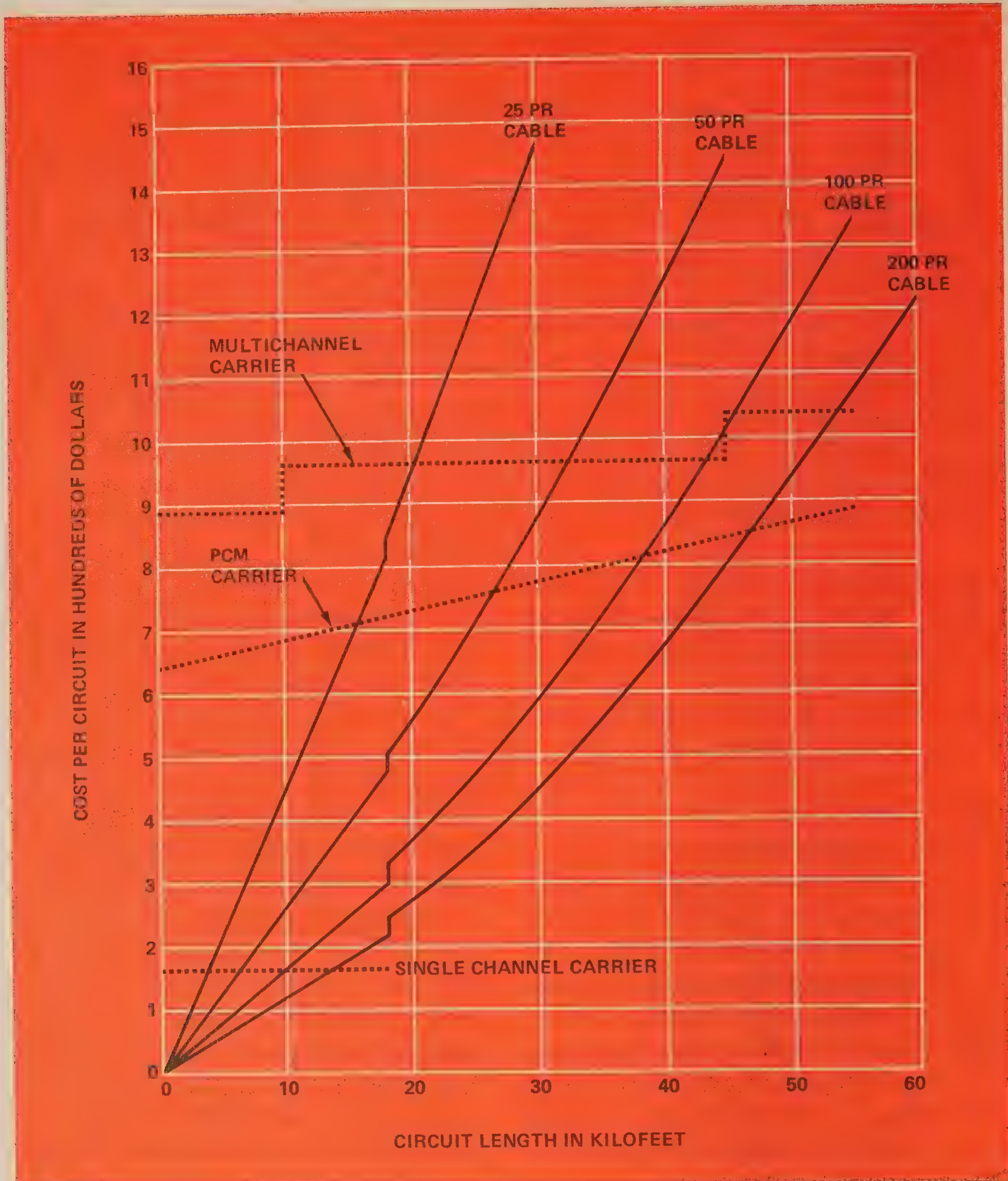


Figure 4. Typical example of comparison of in-place costs for single and multichannel subscriber carrier versus buried cable, with a 1200Ω customer loop-resistance limit. Beyond the crossover point, the in-place cost of carrier is more economical than cable plant.

ary solutions to a problem, to be replaced when additional cable is installed. This same reliability can also make carrier an important component in the design of *new* telephone installations. And, it is no longer limited to providing service to rural areas; carrier is now used in metropolitan areas as well. Planning new cable facilities requires an accurate estimate of the number of cable pairs that will be needed over a period of years. An initial requirement of 50 circuits may increase to 400 circuits in a period of a few years, depending on the population development of an area.

If an accurate estimate of future needs is difficult to make, subscriber carrier can be used to eliminate the risk of providing too many or too few cable pairs. Using a hypothetical situation, an area which requires 50 circuits initially, may require 600 circuits in five years. Rather than install a cable that contains 600 or more circuits, and initially have many idle pairs, a 50-pair cable can be installed. By using PCM subscriber carrier in the initial installation, additional equipment channels may be purchased as needed over an extended period of time. Each PCM group can provide 24 channels on two-cable pairs, and 50 cable pairs can support 25 PCM groups. This means that 600 single-party voice circuits (24 channels X 25 PCM groups) can be derived from 50 cable pairs, if PCM subscriber carrier is used. (This assumes that adequate near-end crosstalk loss exists between pairs to be used for PCM carrier.)

Prove-In Distance

The term *prove-in distance* implies the results of an economic study which compares different mediums; cable and carrier, for example. The point along the length of a facility at which the cost of utilizing one transmission medium or another is equal, is called the crossover point. Before this point or beyond it, one transmission medium is more economical than the other. The distance from the central office at which this point occurs is known as the prove-in distance. Figure 4 is a typical example of how the results of a new-cable installation versus carrier installation comparison might appear. The carriers in this case are of the single-channel and multichannel type, such as the GTE Lenkurt 84A, 82A, and 910A. The prove-in distance for such systems, using a 50-pair cable, is approximately 6 kilofeet for single-channel station carrier, 32 kilofeet for multichannel station carrier, and 26 kilofeet for PCM subscriber carrier. Beyond these distances, carrier would be the more economical choice. Prove-in distances will vary for individual systems, since costs are dictated by such factors as number of cable pairs, gauge size, circuit requirements, and geographical location.

Economics

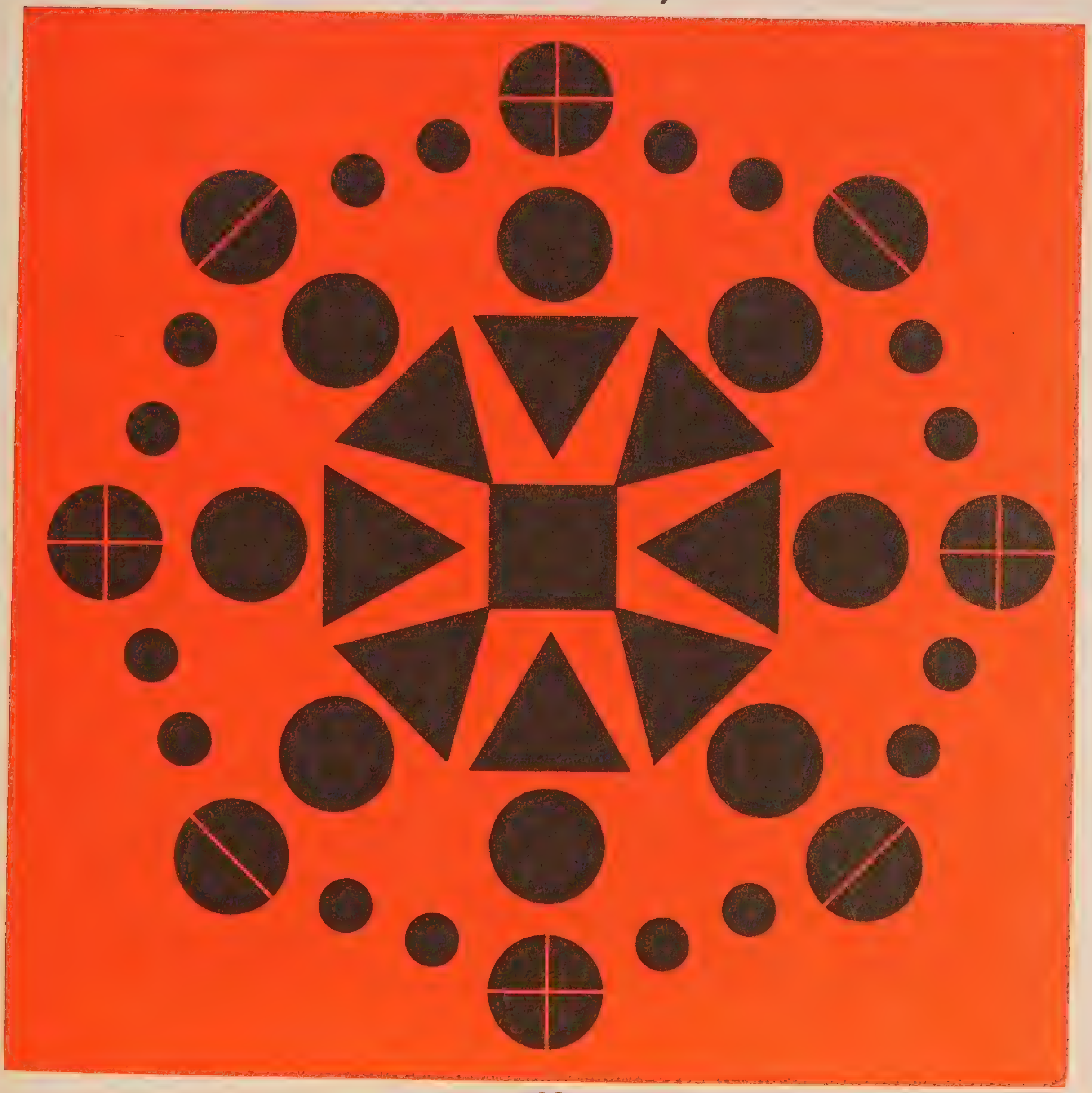
Approximately one third of a telephone company's investment dollar is spent on the customer distribution network. Recent studies based on new cable installation versus carrier installation costs indicate that enormous annual savings could be incurred by the telephone industry by the proper use of these carrier systems – a significant incentive for the proliferation of new carrier in the customer distribution network. The basic reason for this savings is that about 80% of telephone customers are within 18 kilofeet of the central office. This condition is highly conducive to the extensive use of single-channel station carrier systems, since most can operate within the 18-kilofoot distance. PCM carrier also shows great promise, since it can generally prove in at distances as short as 16 kilofeet.

The story of carrier is mainly one of economics. Yet, it also provides technical improvements in transmission, frequency response, and signaling capability. It also eliminates installation of additional cable, loop extenders, voice frequency amplifiers, and long-line adapters. Electronics has played a major role in toll and exchange-trunk facilities for many years. Now, with increasing consistency, electronics will be appearing on the customer's end of the line.

SECTION II
PULSE CODE MODULATION

GTE LENKURT
DEMODULATOR
JANUARY 1972

Toll-Quality PCM



The progress in communications equipment development is everchanging. Some years ago the D1-type PCM terminal made its debut in the communications market. Today, a new generation of toll-quality PCM terminals are making their appearance.

The new generation of PCM (pulse code modulation) terminals for the telephone industry are generally referred to as D2-type terminals. They are so designated because they are end-to-end compatible with the D2 terminal manufactured by the Western Electric Company. The earlier D1-type terminal was designed primarily for short inter-office exchange trunks. The D2-type terminal, such as GTE Lenkurt's 9002A, has been designed for use at any level in the switching network of a country. Both D1- and D2-type PCM terminals operate with 24-channel pulse trains (1.544 megabits/second) and can be transmitted over T1-type repeatered lines.

The telephone switching centers of most countries are classified according to a hierarchial structure. The United States has five different classes of telephone offices. These are numbered in decreasing rank from one to five, a class five office being connected to the subscriber telephone and a class one office being of the highest rank. Figure 1 shows the hierarchical structure of the telephone switching network in the United States. This structure applies to all types of communications equipment, not just to PCM.

Trunks

A trunk is a voice-frequency circuit connecting two telephone switching centers and is therefore used by many subscribers. There are different types of trunks some of which can be classified as follows:

(a) direct trunk – interconnects two class five end offices

(b) toll connecting trunk – connects a class five end office to any higher-ranking toll office

(c) intertoll trunk – connects any class one through four toll switching office to any other class one through four office.

The basic objective of telephone companies is to provide similar voice quality for local as well as long distance calls. Since a long distance call such as an interstate call is usually set up using several intertoll-type trunks connected in tandem, the transmission requirements for such trunks are very strict. By contrast, a local call may involve only one direct trunk between two class five offices. Therefore, the transmission requirements for trunks

Figure 1. Hierarchical structure of the telephone switching network in the United States. Solid lines indicate main lines of connection within the hierarchy; dashed lines show examples of so-called "high-usage trunks" that can be connected between any two offices (regardless of rank) whenever the traffic volume justifies it.

between lower-ranking offices are less severe than between higher-ranking offices. D1-type terminals can be used for trunks between offices of class five up to class three. The D2-type terminal, because of its improved design, is suitable for use at any level in the switching office hierarchy.

The improved design of the D2-type terminal includes improved return loss, delay, and frequency response characteristics as well as patching jacks and precision attenuators for level adjustment. The most important improvement in D2-type terminals is the use of 8-bit encoding which results in reduction of quantizing distortion.

D2-Type Terminals

The Western Electric D2 channel bank (terminal) is equipped to handle a total of 96 channels. It operates with four input and four output pulse trains, each input-output pair connecting to a 24-channel terminal like the GTE Lenkurt 9002A, a similar D2-type terminal, or another 96-channel D2 terminal. Voice signals coming into the PCM terminal are coded into a sequence of pulses and spaces, sent out on a T1 line, and received at the distant end by another PCM terminal which decodes the incoming pulse train and restores the original voice information. This voice information is then sent on to its intended destination by means of telephone lines or other carrier systems. Figure 2 shows a PCM voice communication system using D2 and D2-type terminals.

Quantizing Distortion

The noise in a PCM system is largely due to quantizing distortion which is caused by the approximation inherent in the quantizing process. The input signal to a D2-type PCM terminal is sampled 8000 times per second and each sample is rounded off (quantized) to the nearest of a number of specific voltage levels (steps). Theoretically, the largest number of quantizing steps available in a D2-type terminal is 256 since 256 is the largest number that can be encoded with eight binary digits ($2^8 = 256$). Two different numbers are used to designate the voltage when the sample is quantized to zero amplitude. Since two steps are used for zero voltage, there are 255 steps available for quantization of voltage samples. In order to ensure reliable timing information to the regenerators (repeaters) on the line, suppression of the all zeros code (00000000) is usually required. This leaves a total of 254 quantization levels whenever the all-zeros code needs to be suppressed. Figure 3 gives an example of D2-type quantization at low voltage levels. At the distant end of the system, approximation errors occur because each sample can be decoded only to one of the 254 quantization levels.* These quantization errors appear as distortion.

In contrast to D2-type terminals, the D1-types (which are not end-to-end compatible with the D2-type) use 7-bit encoding which allows 127 quantization levels. This gives a greater distance between steps and consequently, a greater error possibility. The smaller quantizing steps of the D2-type terminal allow a smaller distance between steps which results in less error magnitude and therefore, less quantizing distortion.

Figure 3 shows that steps used in quantizing are of unequal size. This represents the technique used for both D1- and D2-type PCM terminals; the

*There are 254 quantization levels used in five frames out of six. In the remaining frame, 7-bit encoding (resulting in 127 quantization levels) is used, one bit being used for signaling.

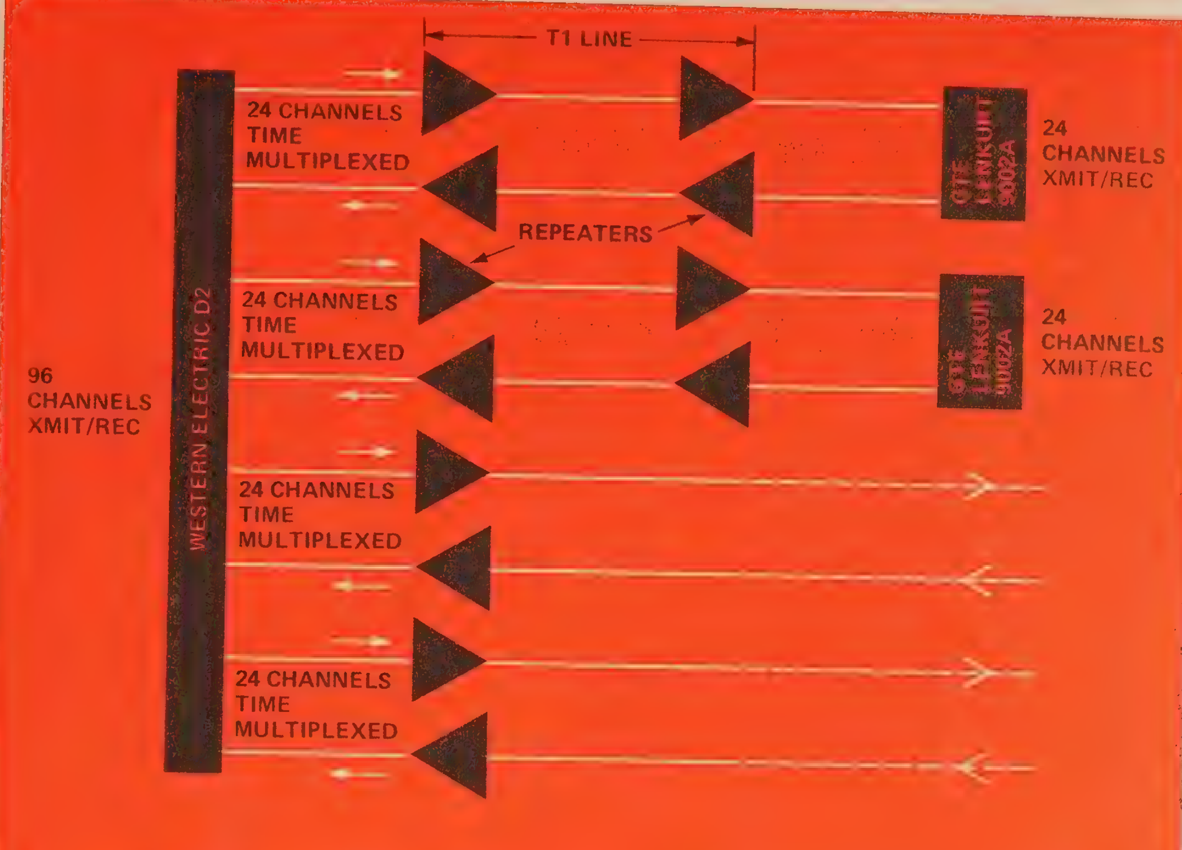
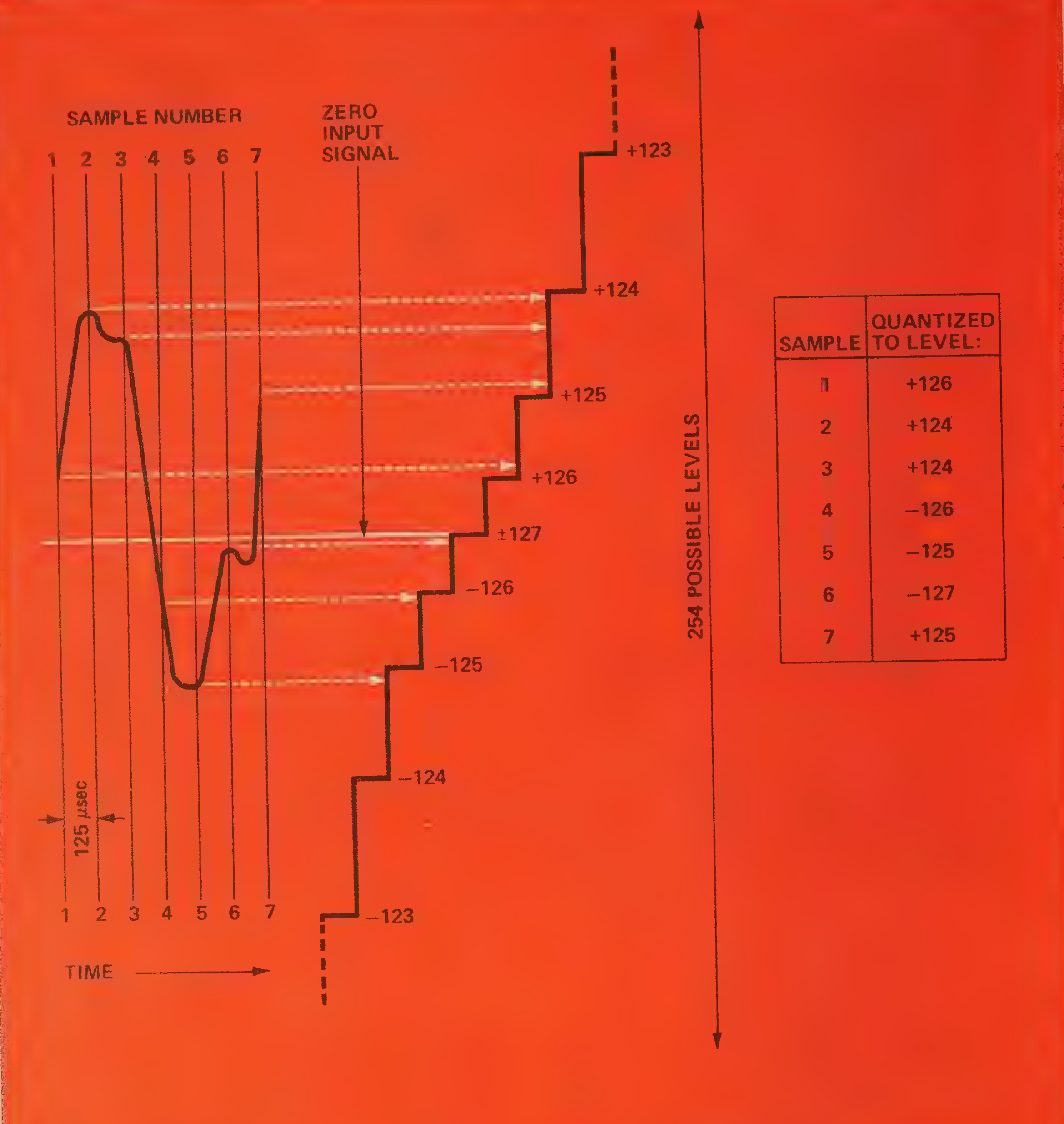


Figure 2. The 96-channel D2 terminal utilizes four input-output signal pairs, each capable of connecting to another D2 or to a D2-type PCM terminal.

steps are always smallest for low power level signals since most speech information is concentrated at low amplitude levels. This results in an improvement in noise performance by virtue of the fact that all signal levels are not equally probable. That is, low voltage signal samples are more likely to occur than high voltage samples in speech signals. The rate at which quantizing steps are allowed to increase is determined by a particular mathematical law especially chosen to optimize the quantizing according to different criteria. For voice communications, the main criterion is a constant signal-to-noise ratio.

Uniform quantizing (equal step sizes throughout the voice range) would introduce distortion with particularly detrimental effects on low amplitude signals. Selection of the most advantageous mathematical law allows the designer of communications equipment to build into his design those properties which are best suited to the required function of the equipment. For example, if signals (such as data signals) of a certain amplitude range and with well-defined statistical properties were to be sent over a communications system, a quantization law could be derived for that system to minimize the error rate. For the transmission of speech signals however, the property desired by telephone companies is a signal-to-noise ratio which is constant over as wide a range as possible. A constant signal-to-noise ratio results in equal quality of



SAMPLE	QUANTIZED TO LEVEL:
1	+126
2	+124
3	+124
4	−126
5	−125
6	−127
7	+125

Figure 3. In a D2-type terminal, each voltage sample is quantized to one of 254 possible levels.

service for weak and loud talkers. A quantized rendering of the lower amplitude speech levels necessitates a specific distinction between discrete levels at low amplitudes, while less specific distinction is necessary at the higher speech levels.

The designation S/D is used here to define the theoretical signal-to-quantizing distortion ratio. Quantizing distortion is generally the chief noise contributor in a properly operating PCM system. The *overall* signal-to-noise ratio (S/N) should be within a few dB of the theoretical optimum represented by the signal-to-quantizing distortion ratio (generally 1 to 5 dB below). That is, the overall signal-to-noise radio (S/N), which includes not only quantizing distortion but also all

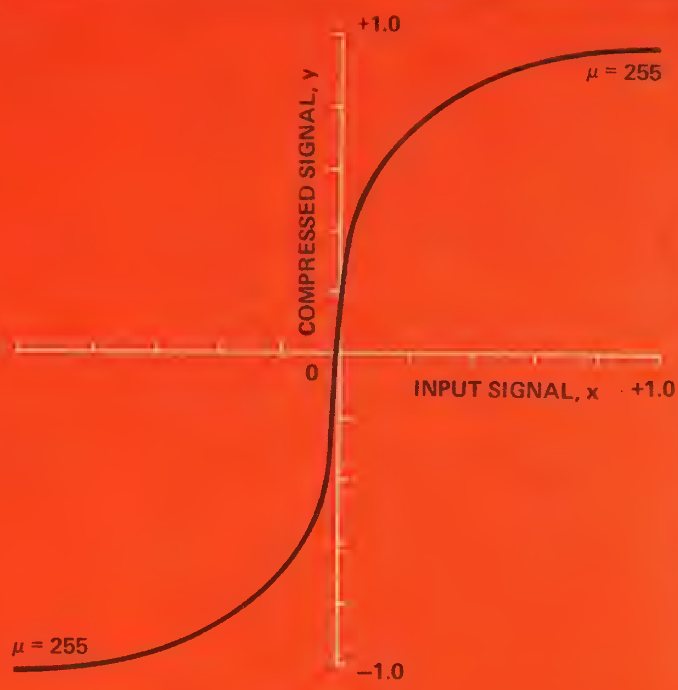


Figure 4. Quantizing characteristic curve used with D2-type terminals for positive and negative values of x, normalized so that maximum input signal corresponds to x=1, maximum output signal to y=1, and μ to 255.

other noise generated within the terminal, should remain within 1 to 5 dB of the theoretical signal-to-quantizing distortion (S/D).

Quantizing Characteristic

The quantizing characteristic determines the rate of increase in size of the quantizing steps. This characteristic is accomplished in the D2-type terminal through the use of a non-linear coder designed to follow the mathematical quantizing law necessary to approximate a constant signal-to-noise ratio. In D1-type equipment, the quantizing characteristic is often called the *companding characteristic* since it is effected through a combination of an instantaneous compressor-expander (compandor) and a linear coder-decoder (equal quantizing steps). In the D2-type terminal, the compression and encoding functions take place simultaneously in the coder although they may be thought of as separate functions, for simplicity.

The quantizing characteristic for a D2-type terminal such as the GTE Lenkurt 9002A follows the equations:

Equation A

$$y = +\frac{\ln(1+\mu x)}{\ln(1+\mu)} \quad \text{for positive x}$$

Equation B

$$y = -\frac{\ln(1-\mu x)}{\ln(1+\mu)} \quad \text{for negative x}$$

where ln stands for natural logarithms (logarithms to the base e, where e is approximately equal to 2.718), x and y are the input and output voltages from the compressing function and μ is a constant which equals 255 for D2-type terminals and which determines the shape of the quantizing characteristic curve. The 255 value of μ should not be confused with the number of quantizing steps as it is a separate entity. The quantizing curve for D2-type terminals is shown in Figure 4. It is approximately linear for

small signal samples, approximately logarithmic for large signal samples (positive or negative), and has a smooth transition between these two limits. In contrast, the D1-type terminal employs a μ value of 100.

Since a logarithmic characteristic is desirable for a constant signal-to-quantizing distortion (S/D), the value of μ = 255 implies an improvement over the D1-type equipment because the characteristic assumes an approximately logarithmic shape at smaller values of input signal x. The net effect is that S/D stays approximately constant over a wider range of input signal values, as shown in Figure 5.

Although quantizing distortion is the predominant noise source in a properly operating PCM system, there are other sources of noise. In the analog section of the terminal, noise from harmonic distortion as well as crosstalk between channels can occur. This is particularly true if several channels are loaded with high power level signals. Errors may also occur in the sampling process (before quantizing) and during encoding and decoding of the signal, as for example, when there is a slight error in dc level in the encoder. All of these factors cause the overall signal-to-noise ratios to be somewhat less than the theoretical signal-to-quantizing distortion ratios depicted in Figure 5.

Signal-to-Noise Ratios

The reader who is mainly acquainted with FDM (frequency division multiplex) may wonder why the signal-to-quantizing distortion ratios shown in Figure 5 are relatively small compared with those commonly expected from an FDM system. For example, FDM cable carrier systems have signal-to-noise ratios usually in the range of 55 to 65 dB, and on a single microwave hop, S/N performance is often 75 dB or better. The D2-type curve in Figure 5 shows that the S/D ratio never exceeds 38 dB, C-message weighted, but this does not mean that PCM systems operate at a great dB disadvantage with respect to FDM. The difference lies in that Figure 5 depicts the difference in dB between the *actual* instantaneous level of a sine-wave signal and the quantizing distortion level. In other words, the quantizing distortion varies with the signal; a signal level of 0 dBm0 and a signal level of

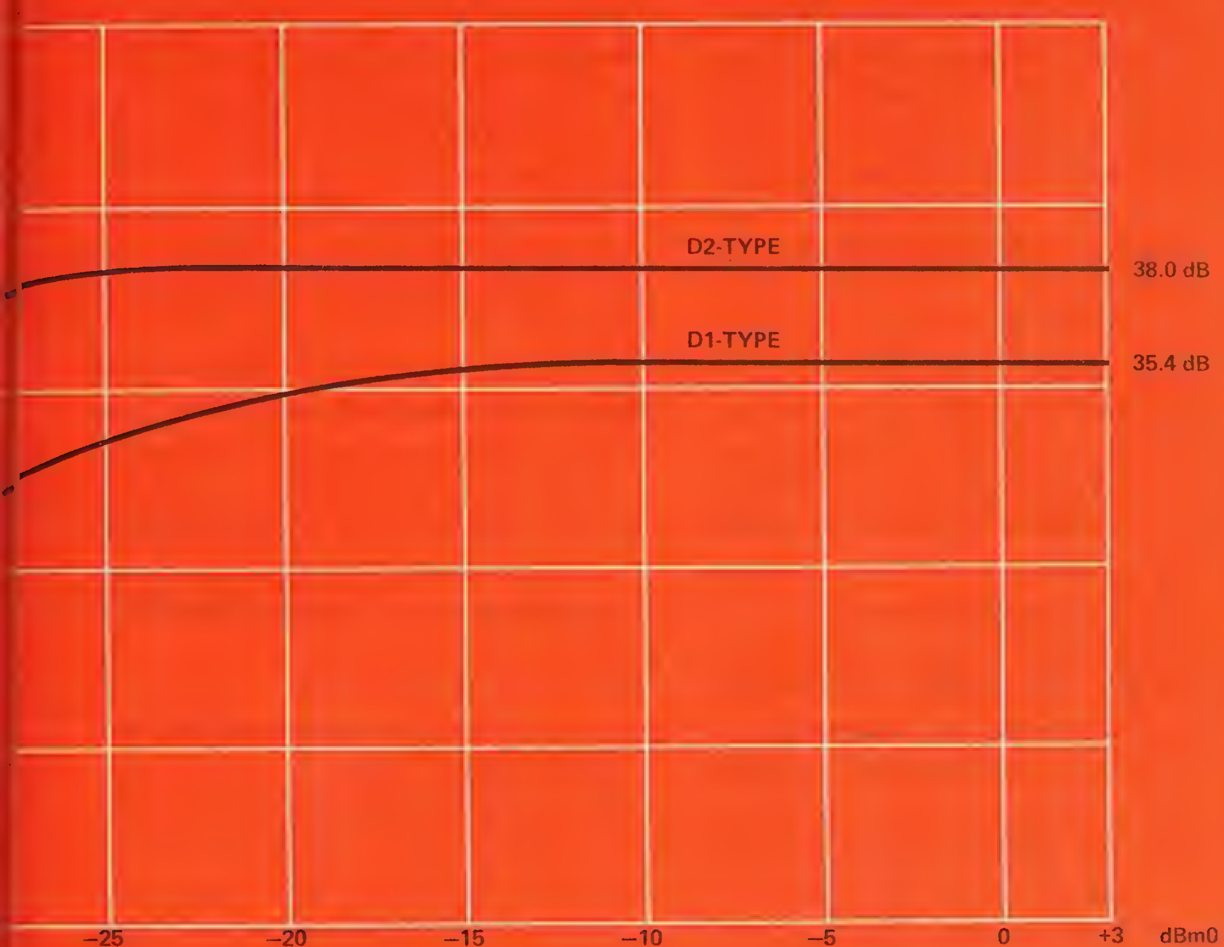


Figure 5. Theoretical signal-to-quantizing distortion ratios (S/D) for D2- and D1-type PCM terminals.

–30 dBm0, looked at separately, result in entirely different levels of distortion (the distortion varies with the signal in such a way that the S/D ratio remains approximately constant between +3 and –30 dBm0). The dB values obtained from Figure 5 are thus representative of the amount of quantizing distortion a talker will actually experience when talking over the system; S/D ratios of 20-30 dB are usually quite adequate.

In contrast, "S/N" in an FDM system nearly always means "test tone-to-noise" ratio *referenced to a test tone level of zero dBm (1 milliwatt) at the zero level point.* The noise in an FDM system is present all the time, whether there is a signal or not. It is only to a small extent dependent on signal level; the noise is practically constant over a wide range of signal levels. In such a system, a S/N performance of 65 dB means that the difference in level between test tone level (inserted with zero dBm level at the zero level point) and noise level is 65 dB. If a weak talker uses this FDM

system at a signal level of –45 dBm0 (45 dB weaker than test tone) he will experience an "actual S/N ratio" of only 20 dB (65 dB – 45 dB = 20 dB). It is the obligation to provide satisfactory service to weak as well as loud talkers that requires provision of such wide margins in FDM systems.

Idle Noise and Interchannel Crosstalk

Improvement in signal-to-noise ratios at low signal levels is not the only advantage of a high value of μ such as 255 (as compared to μ = 100). Another advantage of a large μ value is a reduction in idle noise and interchannel crosstalk due to irregular excitation of weak-signal quantum steps.

The idle circuit noise – as well as the crosstalk into an idle channel from a channel with traffic – depends on several factors, but mainly on the quantization step sizes that correspond to very small voltage samples. One reason for this is due to the tolerance in the sampling process. Slight imperfections in the sampling circuitry may cause a small dc offset voltage at the input to the coder. This offset voltage may be large enough that a very small amount of noise or crosstalk can trigger the next (positive or negative) quantization step voltage. This produces an increase in noise or crosstalk. The magnitude of this increase is smaller when the steps are smaller. Idle circuit noise performance as well as interchannel crosstalk performance is improved for D2-type systems because of new circuit design, increased number of steps, and greater μ value (resulting in smaller step sizes) compared to the D1-type terminals.

Seeing that a higher value of μ gives certain advantages, why isn't an even higher value of μ (than μ = 255) selected? There are basically two reasons for this; one is that end-to-end compatibility with other US-manufactured PCM systems would not be possible and the other is that difficulty would arise in designing the appropriate equipment without excessive cost, still retaining the added advantage of a higher μ value.

There are several difficulties in implementing high μ values without adverse results. Some of these difficulties include:

(a) adhering to necessary tolerances while maintaining the reciprocal relationship between the non-linear characteristics of coder and decoder characteristics (between compressor and expander curves in D1-type terminals)

(b) keeping the dc component of the multiplexed signals to a low enough value so as not to offset the advantage of a high μ. To prevent these phenomena from canceling out the advantages gained by a high μ value, smaller, more costly, design tolerances are necessary.

Implementation of the Quantizing Characteristic

In a D1-type PCM terminal, the quantizing characteristic used has a value of μ = 100. This characteristic is achieved through the use of a diode compressor in the transmitter, followed by a uniform coder (giving equal quantizing steps). The end result of these compressing and coding processes is the same as if the quantizing had been implemented directly using unequal steps. The temperature sensitivity of diodes requires that they be mounted in an oven resulting in an increase of total power consumption.

In a D2-type terminal, the quantizing characteristic can be imple-

mented in a non-linear coder through the use of a so-called "piecewise" linear coding process. The desired quantizing characteristic using μ = 255, is approximated by fifteen connected straight lines (covering fifteen different regions of input signal) of varied slopes, the slope of each line segment determining the step sizes in that particular range of input signals. More simply, the quantization curve, shown in Figure 4, and developed from equations A and B, is made up of 15 different connected lines, each with a particular slope. These 15 lines extend from the largest negative to the largest positive signal input voltage which can be sampled by a D2-type PCM terminal without excessive distortion. The result is better accuracy and less power consumption than when diode implementation is utilized.

CCITT and International Standards

The CCITT (International Telegraph and Telephone Consultative Committee) has adopted standards for PCM systems, distinguishing between so-called "primary" and "secondary" PCM multiplex equipment. Primary equipment multiplexes voice or data signals into time-multiplexed sequences of pulses and spaces, whereas a secondary multiplexer accepts several primary PCM pulse trains and multiplexes them into PCM pulse trains with a higher bit rate. Secondary multiplexers have been under development for some time but have only been put into limited service to date.

It would be desirable, from the point of compatibility, if there existed a single world-wide standard for both primary and secondary multiplexers. Although this has not as yet been possible, a step in that direction was taken when the CCITT agreed to recommend two types of primary systems as standards. The two systems standardized were a D2-type system and a European system for 30/32 channels (30 voice channels plus 2 channels for time-shared signaling) using a higher bit rate than the D2-type system. Whether further steps in world-wide standardization will be long or short in coming, only the future can tell. Meanwhile, the D2-type terminal exists as an improvement in the everchanging field of communications.

GTE LENKURT
DEMODULATOR

MARCH 1971

PCM Signaling and Timing

The integration of pulse code modulation (PCM) carrier systems into the telephone plant has left personnel, expert in dealing with the traditional frequency-division multiplex systems, grappling with less familiar terms such as "sampling," "quantizing," and "coding."

The principles of message transmission in a PCM system have been described in a variety of articles and books. (See, for example, *The Lenkurt Demodulator*, November, 1966.) While message transmission is, of course, the objective of such systems, it is not the whole story. A message with no place to go is no message at all.

How, then, do PCM systems carry the various types of signaling and supervisory information that control an ordinary telephone call? And how can a PCM carrier system integrate into an existing plant that already includes such diverse types of signaling as E&M (receive and transmit), loop dial, and foreign exchange? Furthermore, how are these signaling and supervisory functions handled in second generation PCM systems?

The answers to these questions are inextricably bound up with the carrier's nervous system – the timing arrangement that sorts out more than a million-and-a-half information bits each second to form individual message channels and their associated signaling information. A good starting place is a brief review of the transmission techniques used in first-generation PCM systems. These systems are built by several manufacturers. Regardless of their origin, however, they conform to the same general system parameters.

T-Carrier Transmission

For convenience, the entire carrier system is referred to here as a *T-carrier* system, in accordance with the Western Electric Company designation. However, common usage has separated the *T1 repeatered line* from the *D1 channel bank* – the actual multiplex terminal. It is in the D1 bank that sampling, quantizing, and encoding occur. It is also the D1 bank that controls system timing, the all-important brain of the system.

The analog voice signals are first sampled in sequence to form pulse amplitude modulated signals. Each pulse is then quantized – assigned the nearest discrete value to its actual amplitude. Logic circuitry then encodes the pulse into a binary number that defines this discrete value. This binary number is expressed as a series of identical pulses, or spaces. A pulse indicates a binary "1" and a space indicates a binary "0".

The series of pulses and spaces that defines one quantized sample from one channel makes up a PCM word.

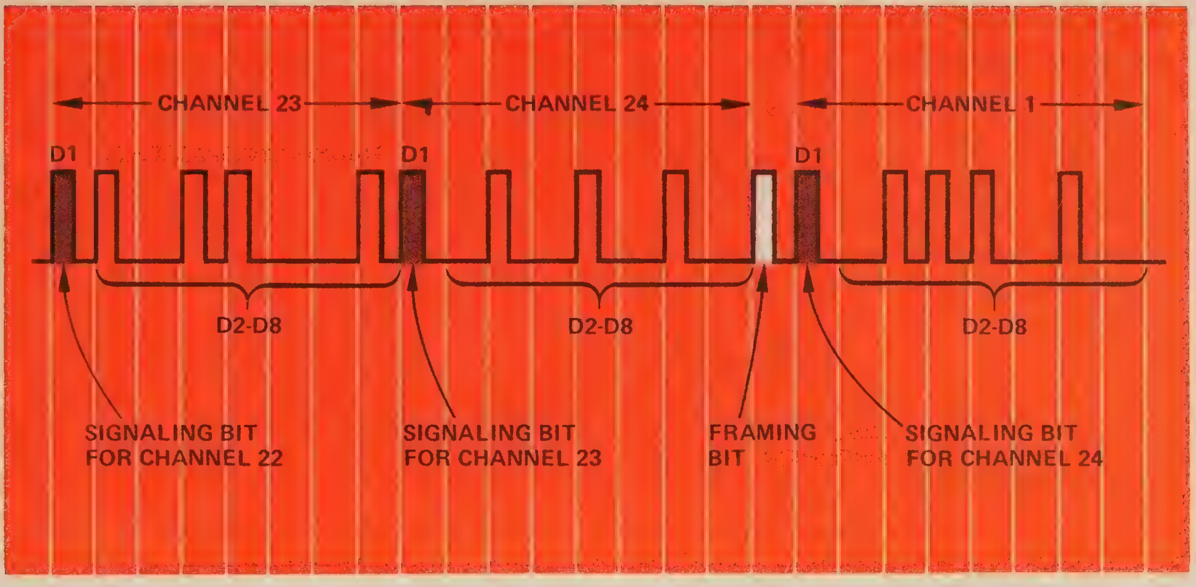


Figure 1. Frame format of D1 channel bank consists of 24 8-bit PCM words plus one framing time slot. The D1 time slot in each word is reserved for signaling information for the previous channel.

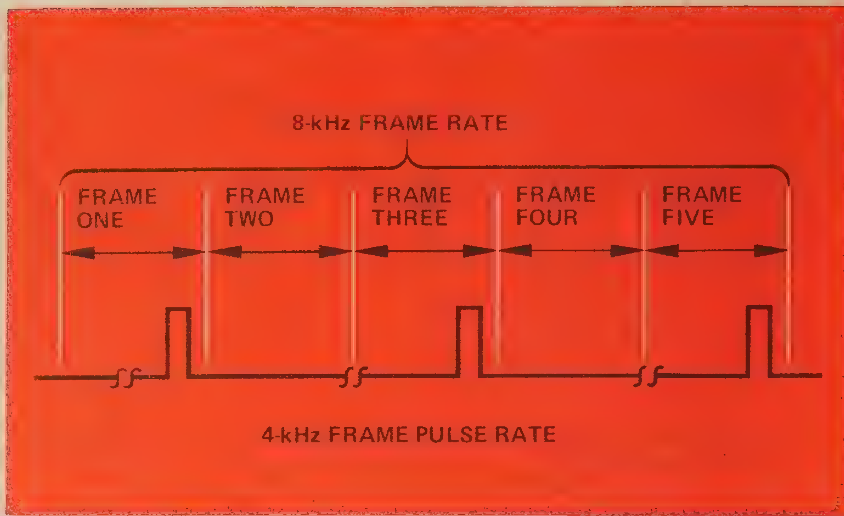
The length of the word limits the number of quantizing steps that can be used, and hence the fidelity with which the original analog signal can be reproduced at the receiving terminal. The D1 bank uses a seven-bit encoding scheme, which permits 2^7, or 128, quantizing steps. (Other considerations preclude the use of 0000000, so 127 steps are actually available for quantizing the voice signal.)

However, there is one necessary ingredient in the PCM word associated with one sample from a single channel. This is some way to carry the signaling and supervisory information. In the D1 bank, this is done by adding one additional bit to form an eight-bit word. The first, or D1, time slot in each word is reserved for signaling and supervisory information for the previous channel.

Since the system handles 24 voice channels, 24 eight-bit words (a total of 192 bits) are contained in one scanning cycle – one word from each channel. These 192 bits make up a "frame." Without a means for the receiving terminal to identify the beginning and end of these frames, the transmitting and receiving terminals will not be synchronized and the receiving terminal will be unable to route the individual words to the appropriate channels. Therefore, a 193rd time slot is inserted in each frame, as shown in Figure 1, to provide timing information.

This framing bit alternates between a "1" and a "0" for succeeding frames. The result is a stable signal component at one-half the frame rate. Since the frames recur at a rate of 8-kHz, the alternating framing pulses produce a 4-kHz component, as shown in Figure 2. The framing circuitry in the receiver locks onto the frame rate. In the event of loss of synchronization, the receiver "slips" one bit per frame until it regains synchronization. If it has not regained the frame rate after checking each bit in two frames, an alarm is initiated. It takes 48.25 milliseconds to check these 386 bits.

Figure 2. Alternating one and zero framing bits produce a 4-kHz pulse rate to establish the synchronization between transmit and receive terminals.



Transmitting Signaling Information

As far as the D1 bank is concerned, there are two types of nonmessage information to be transmitted: supervisory information (on-hook, off-hook) and signaling information (dial pulses and multi-frequency tones). Supervisory information is transmitted using two possible electrical states such as – open or closed loop; potential on either side of the incoming line; or battery or absence of battery on the signaling leads. These two varying electrical states result in a series of either 1's or 0's in the D1 time slot of, say, channel one. When the electrical state changes, the 1's change to 0's, and vice versa. At the receiving terminal, the series of pulses and spaces is used to reconstruct the original DC potential for transmission to the office switching equipment.

Transmission of dial pulses is nearly as simple as transmitting the supervisory information. Assuming, for ease of calculation, a 50/50 make/break percentage, a dial pulse at a nominal 10-pulse-per-second rate has a duration of 50,000 microseconds. Since a sample is taken every 125 microseconds (in the original D1 bank), each pulse is sampled 400 times. (See Figure 3.) Thus, neither the pulse rate nor the make/break ratio is critical. The PCM system sees dial pulses as slowly changing potentials.

Since there are only two possible states in both supervisory information and dial pulses, neither needs to go through the quantizing process used for voice signals. All that is required is sampling at the appropriate time and conversion to the correct voltage level at the receive terminal. Thus, the signaling and supervisory information enters the transmission path just before the bit stream goes on the line. Conversely, this information is extracted from the bit stream as soon as it comes off the line at the receive terminal.

Multi-frequency signaling tones consist of varying AC within the voice band. Therefore, the D1 bank treats them like voice signals. It samples, quantizes, and encodes them. At the receive terminal, they are reconstructed in the same manner as a voice signal. Thus, when multi-frequency signaling is used, the actual signaling path in the carrier system handles only supervisory information.

There are two possible separate signaling paths through the entire common carrier equipment. Not all signal-

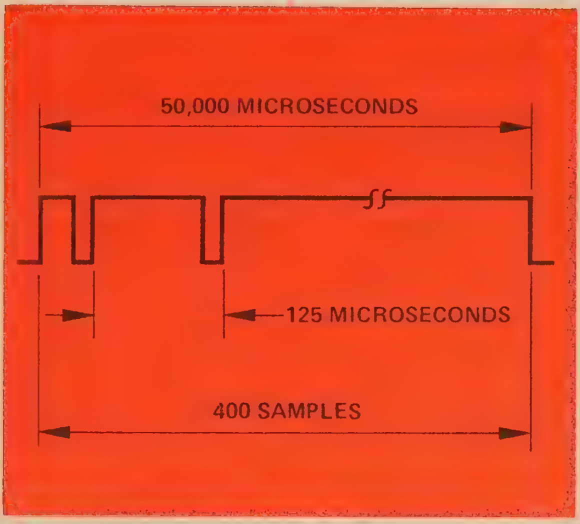


Figure 3. Each dial pulse is sampled approximately 400 times at an 8-kHz sampling rate.

ing arrangements require both paths. Dial pulse and E&M signaling, for instance, each need only one path. However, more complex signaling schemes that must send two types of information simultaneously require both paths.

For example, foreign exchange signaling arranged for forward disconnect must hold the subscriber terminal busy while the office disconnects. It is a matter of controlling two relays, one to hold the subscriber off-hook, and the other to provide on-hook/off-hook information about the office condition. Nevertheless, two separate signaling paths are required for such an arrangement.

Two Paths on One Bit?

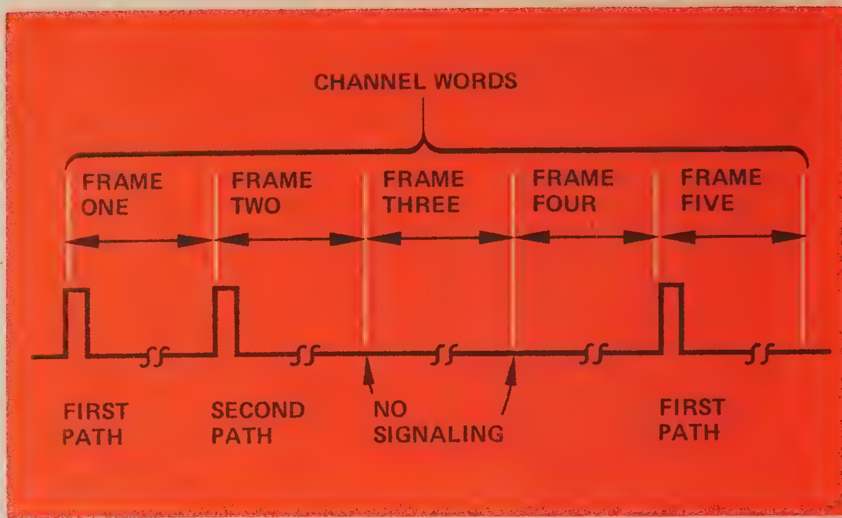
Since two signaling paths are necessary for certain types of signaling, and only one bit in each word is set aside for signaling information, how can the two paths be kept separate? Western Electric Company has developed two separate signaling schemes for the D1 channel bank. These two schemes are called *D1A* and *D1B.*

Although only one digit is set aside for signaling, it is possible to borrow one of the voice digits to provide the second signaling path. The D1A arrangement borrows the eighth bit of the PCM word (the least significant bit) to provide the second signaling path. Hence, this option is often referred to as D1/D8 signaling. Even though this technique uses one of the voice digits, it does not affect the quality of voice transmission through the channel; since once the call is established, D8 is returned for exclusive use in voice encoding. A pulse in D1 indicates the channel is in an on-hook condition. When no pulse appears in the D1 time slot, the called terminal has gone off-hook, and message traffic is imminent. The absence of a D1 pulse inhibits the use of the D8 time slot for signaling, freeing the D8 time slot for full seven-bit voice signal encoding.

This arrangement works out well except in cases where the called terminal sends back no on-hook/off-hook supervisory information. These are the so-called "free" calls (to directory assistance or a test line, for example) where there is no reverse battery supervision. In such a case, a pulse appears in the D1 time slot even when the called terminal goes off-hook. Therefore, D8 would continue to be used for signaling, using a digit that would normally be reserved for voice transmission. As a result, the voice signal is encoded in only six bits instead of the usual seven. The increased quantizing noise with 63 quantizing steps, rather than the usual 127, substantially degrades the quality of the voice channel.

This condition is not a universal problem because it only occurs with certain types of signaling — and then only on free calls. Nevertheless, it can

Figure 4. In D1B signaling, the first signaling path is provided by the D1 time slot on first frame, the second signaling path uses the D1 time slot of the second frame, and neither uses it on third and fourth frames.



be avoided with a different technique for providing two signaling paths. This improved two-path arrangement is referred to as D1B.

Since D1B uses only the D1 time slot for signaling, it is sometimes called "D1 only." The D1 time slot in the first frame provides the first signaling path, the D1 time slot in the second frame provides the second path, and both paths are inhibited during the third and fourth frames. Then the pattern repeats. This four-frame pattern, shown in Figure 4, is necessary to avoid confusing the receiving terminal with false framing information. Suppose, for instance, that the signaling paths were to use the D1 time slot in alternate frames, and one of them produced a series of pulses while the other produced no pulses. The resulting series of alternating 1's and 0's would produce a 4-kHz fundamental component that would be indistinguishable from the framing bits.

Each signaling path for a particular channel is sampled only once every four frames and the entire frame length is 125 microseconds; therefore, samples of signaling information are taken every 500 microseconds – or about 100 samples during a typical dial pulse.

Second-Generation PCM Systems

The second-generation PCM carrier terminal is the *D2 channel bank*. Like the D1 bank, D2 uses an eight-bit PCM word. However, the D2 bank is intended to meet intertoll requirements for lengths up to 500 miles. Seven-bit encoding is not good enough to achieve this objective. The quantizing noise would be too high. Therefore, it is not possible to reserve one digit out of the eight to provide signaling information.

The solution is a second level of time-division multiplexing. In five out of every six frames, the D2 bank encodes the voice signal in eight bits. In the sixth frame, it uses only seven-bit encoding, borrowing the eighth bit for signaling information. The result is an average 7-5/6 bit encoding for the voice signal. This improved performance meets the intertoll objectives.

Two signaling paths, for four-state signaling, are provided in much the same way as in the D1B channel bank. The signaling bit in every other sixth frame carries one two-state channel, while the same bit in alternate sixth frames carries the other two-state channel. In this way, complete information about the condition of both signaling paths can be transmitted in

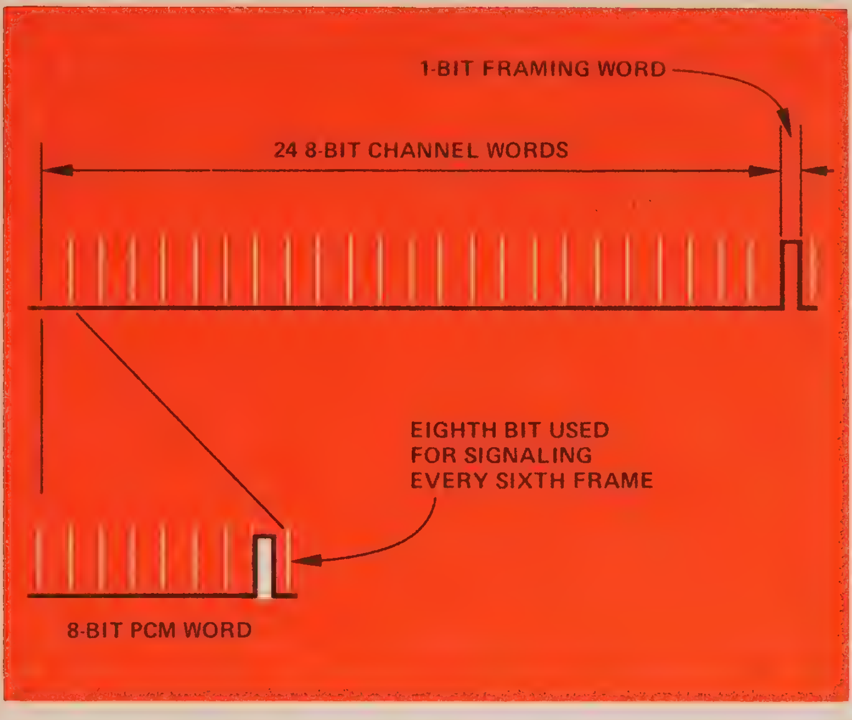


Figure 5. D2 frame format consists of 24 8-bit words plus a 1-bit framing word. Signaling borrows one bit from each channel word, in every sixth frame.

12 frames – about 1.5 milliseconds. If only one signaling path is required, as in the case of E&M signaling, both paths are still used – providing signaling every six frames.

One effect of this time sharing every sixth frame is the necessity for more framing information in the eighth bit of each word. Not only must the receiving terminal recognize the beginning and end of each frame, but it must also determine whether or not a particular frame carries message or signaling information on the eighth bit of each word. Once again the answer is time sharing. The framing bit in every other frame contains the information for terminal synchronization (see Figure 5). This leaves the framing bit in alternate frames free to carry the information necessary to distinguish the one frame in six that carries signaling information.

The net result is a gross frame format and an operating bit rate identical to that of the D1 bank. However, the D1 and D2 banks cannot be operated end-to-end. Not only do their framing and signaling arrangements differ, but they also have different PCM coding schemes and companding characteristics.

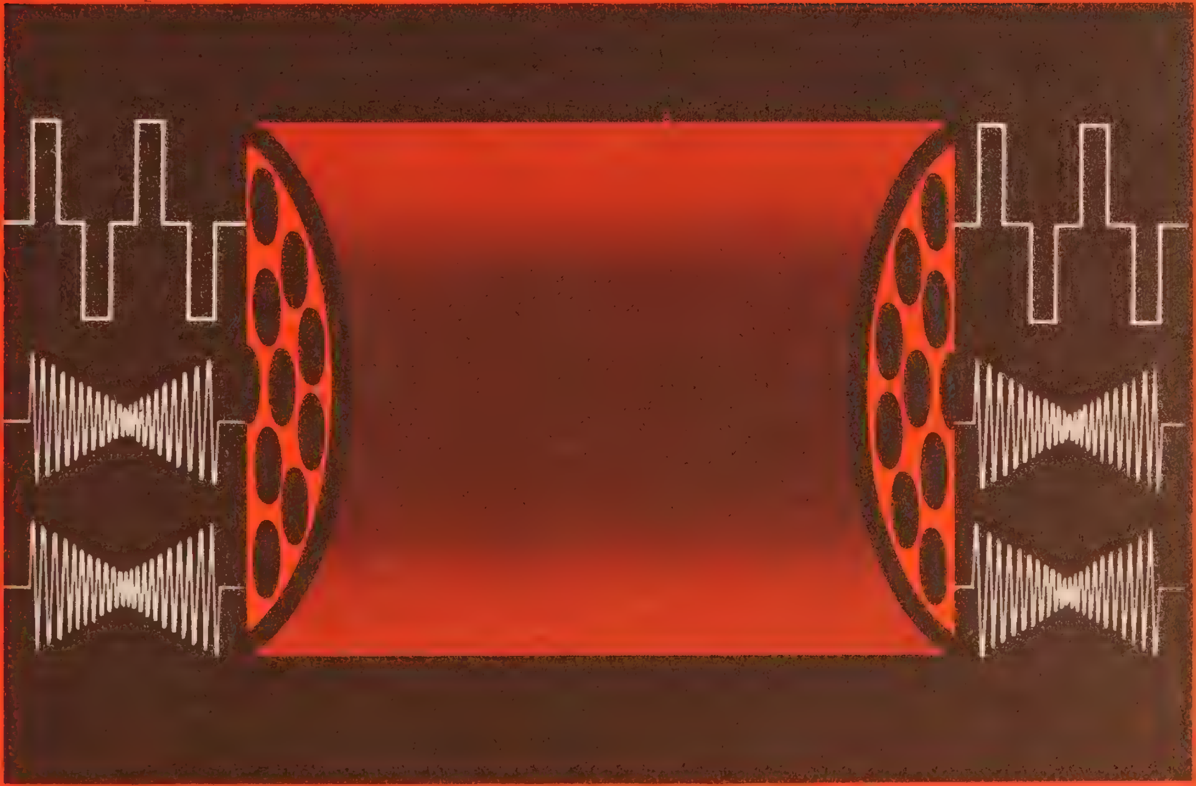
While these two channel banks – D1 and D2 – lack such direct compatibility, they can operate over the same repeatered lines. And they are closely related members of the emerging family of digital transmission systems.

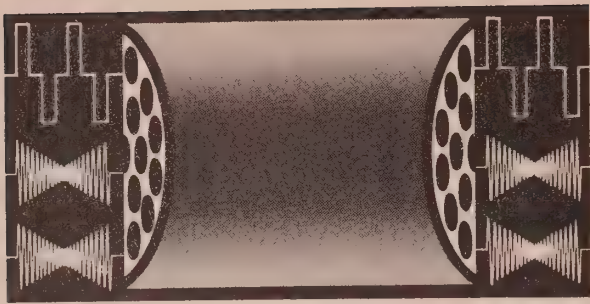
GTE LENKURT

DEMODULATOR

JULY 1971

PCM-FDM Compatibility Part 1





Under certain conditions, pulse code modulation cable carrier systems are compatible with frequency division multiplex systems on cable pairs within the same cable sheath.

It has been the practice in the communications industry to avoid placing pulse code modulation (PCM) and frequency division multiplex (FDM) systems in the same cable sheath. The reason for this taboo stems from the fact that the signal level on a PCM system is so high in comparison to that of an FDM signal that indiscriminate mixing of the two systems may result in rendering the FDM system partially or totally useless. Nearly all noise and crosstalk between the two systems is unidirectional, from the PCM system to the FDM system. The performance of a PCM system will hardly ever be affected by the presence of an FDM system on the same cable.

Because there are economic as well as convenience advantages to the user in combining PCM and FDM systems within the same cable sheath (whether for a temporary or extended period), tests have recently been conducted at the GTE Lenkurt laboratories on the problem of PCM-FDM compatibility. The preliminary results of these tests and the tentative ground rules that have been subsequently established may serve as a guideline to the user who is comptemplating a combination of PCM and FDM systems in the same cable sheath.

Why Combine PCM and FDM?

There are several instances when a user of cable carrier systems may desire to operate PCM and FDM systems over cable pairs within the same cable sheath. For example, he may want to gradually phase out an existing FDM cable carrier system and convert to PCM circuits over a cable route. Or, he may want to install PCM systems on a cable which already has FDM carrier in it to avoid the expense of installing new cable. The FDM carrier, in this case, may include subscriber carrier, N-type carrier or exchange carrier systems.

There may be many cable pairs in one cable and although the metal sheath that encompasses them provides protection from external interference, interference generated within the cable by some of these pairs may cause noise which degrades the quality of the intelligence being conveyed on other pairs.

It is important to realize that the PCM line signal of all 24-channel, T1-type, U.S.-manufactured PCM systems using D1-type channel banks look identical. Also, all D2/T1-type line signals look identical to each other (even though they are slightly different than D1/T1-type signals). Therefore, if one manufacturer can mix PCM and FDM systems on a cable, any other can also. The references to D1- and D2-type systems in this discussion imply particular types of 24-channel PCM terminals (channel banks) both of which are often used in combination with the T1-type repeatered line (regenerative repeaters and associated equipment).

PCM-FDM Possibilities

Whether it is possible to operate a PCM system in the same cable sheath with one or several FDM systems is determined by the crosstalk coupling loss between the cable pairs involved. Given a certain crosstalk coupling loss in a cable between the FDM cable pairs and the PCM cable pairs, compatibility then depends on the frequency range of the FDM system, the baseband frequency of the FDM channels equipped, the number of PCM systems involved, the length of exposure (in terms of number of FDM repeater sections), kind of FDM system (modulation method), and the permissible amount of performance degradation allowed in the FDM system due to PCM carrier interference. When all of these factors are considered, they are evaluated against the power distribution in a T1-type PCM signal as a function of frequency (the power spectrum). What that power spectrum looks like is of vital importance when PCM-FDM compatibility is evaluated. The highest frequency slot in the FDM system will always be the one of most concern, since it will be the slot most vulnerable to interference from the PCM signal.

As a general rule, interference into the FDM carrier system can be reduced by equipping only the lower-frequency channels in the FDM system. Therefore, when phasing out an FDM system which must operate for some time on the same cable sheath with PCM, the higher FDM channels should be phased out first. The power spectrum components in a PCM line-signal drop sharply below 96.5 kHz, and interference into FDM systems below that frequency is usually negligible. The point of maximum power in a T1-type PCM power spectrum occurs at approximately 710 kHz for busy hour conditions. The maximum PCM power, in this case, indicates the maximum amount of interference that threatens the FDM system.

Cable Characteristics

The achievable crosstalk coupling loss between two cable pairs increases with the number of cable pairs in the cable sheath. This is because there is increasingly less crosstalk coupling in proportion to the physical distance between pairs.

The actual value of crosstalk coupling loss between two cable pairs depends upon which splicing groups have been selected, the splicing methods used, and the general crosstalk characteristics of the cable (such as cable gauge and dielectric insulation material).

Direction Coordination

A PCM repeatered line laid out for one-cable operation has a minimum near-end crosstalk (NEXT) coupling loss requirement between its two directions of transmission. Failure to meet this requirement may result in interference between the two directions. Such systems are often planned with the pairs for opposite directions of transmission assigned to nonadjacent splicing groups (see Figure 1). It is therefore important to keep the FDM pairs protected against interference from either direction of transmission of the PCM system. However, PCM to FDM interference between cable pairs belonging to the same direction of transmission is not nearly as serious as interference between pairs belonging to opposite directions of transmission. The reason for this is that the difference between the signal level on a PCM cable pair and that of an FDM signal on another pair, is at most points along a cable, greater for opposite directions of transmission than between pairs for the same transmission direction. The crosstalk disadvantage is thus greatest between cable

pairs for opposite directions of transmission. Figures 2A through 2C show the near-end, and far-end crosstalk (NEXT and FEXT) characteristics for different directions of transmission in various systems. It is assumed that all FDM system repeaters coincide with repeater locations of the PCM system. One repeater section of the FDM system may correspond to one or several repeater sections of the T1-type PCM system.

The only NEXT paths of significance between the T1 and N3 carrier systems shown in Figure 2C are the ones indicating near-end crosstalk over repeater section 3. The contributions from the other two repeater sections arrive at the FDM repeater greatly attenuated and can be neglected. Therefore, this cause can be treated as if all the near-end crosstalk on this FDM repeater section originated on the T1-type repeater section adjacent to the FDM repeater input.

Likewise, FEXT coupling between T1-type carrier and N carrier is due

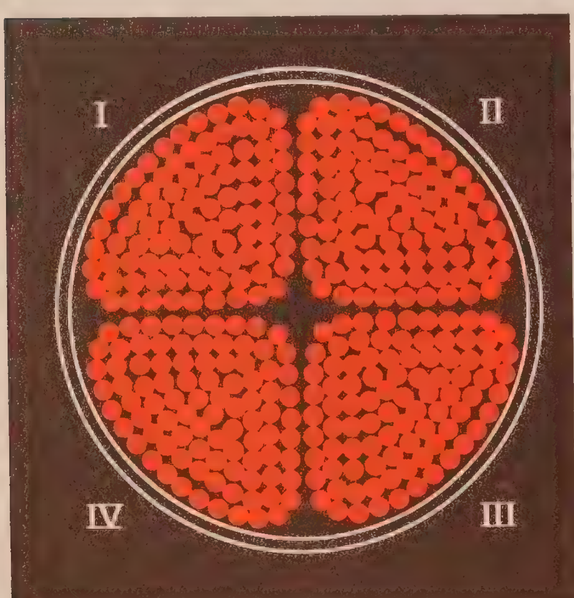


Figure 1. A cable divided into four binder groups. At a splice, if group integrity is maintained, I and III, II and IV, are still non-adjacent beyond the splice. Also, I and II, IV and III, are still adjacent beyond the splice.

almost totally to the FEXT coupling over the T1-type repeater section adjacent to the FDM repeater input.

If a PCM system shares a cable with an FDM system for a distance comprising more than one FDM system repeater section, the interference will add up on a 10 log k basis, where k is the number of FDM system repeater sections exposed to PCM interference.

The number of interfering PCM systems also has an influence on the determination of required crosstalk coupling loss, since the noise powers add up. If the number of PCM systems is n, and 10 log n is used to account for the number of systems, the assumption is then made (conservatively) that the PCM disturbers all interfere with the FDM system at equal coupling losses.

Each PCM system engineered for one-cable operation interferes with each FDM system both by way of near-end and far-end crosstalk. Since such a cable system has the pairs for both transmission directions inside the same cable sheath, the interference into each direction of transmission of an FDM system thus originates from both transmission directions of each PCM system (two cable operation implies that both directions of transmission are assigned to pairs contained within separate cable sheaths).

Considering that a one-cable PCM system should be engineered with pairs for opposite directions of transmission in different binder groups (often non-adjacent in the cable), in some cases one type of crosstalk (near-end or far-end) may dominate over the other. In other cases, it may be necessary to conservatively assign half of the crosstalk contribution to each type of crosstalk, when estimating minimum crosstalk coupling loss between PCM and FDM cable pairs.

The ideal condition is when the FDM systems are assigned to pairs in a

splicing group or unit in the cable which lie nonadjacent to any of the two groups or units used for the two directions of transmission for the PCM carrier. If this is not possible, direction-coordination should then be considered. This implies that the two directions of transmission of the FDM system be coordinated with the PCM carrier pairs in such a way that pairs belonging to the same directions of transmission for the two types of systems are assigned to the same splicing group or unit in the cable.

Screen Separation

Several manufacturers of multi-pair cable have developed an internal screen which allows separation of cable pairs into two compartments. This screen is intended to provide electrical partitioning between pairs used for opposite directions of transmission, thus reducing near-end crosstalk in PCM systems. This offers an opportunity for direction coordination, if FDM systems are to be transmitted over such cable along with PCM systems.

Figure 2A. Near-end crosstalk (NEXT) from T1-type carrier to X carrier (these systems utilize repeater sections of approximately equal maximum length).

Figure 2B. Far-end crosstalk (FEXT) from T1-type carrier to X carrier.

Figure 2C. Near-end crosstalk (NEXT) from T1-type carrier to N3 (Lenkurt 46B) carrier (N carrier utilizes approximately three times longer repeater spacings than T1-type carrier.

It has been suggested by one manufacturer of screened cable that the screening concept may be found useful for providing isolation between PCM and FDM systems. For example, two screens could be provided, dividing the cable core into three compartments, two for PCM and one for FDM usage.

The PCM Power Spectrum

Aside from the observation of direction-coordination between cable pairs, other important factors must be taken into consideration when investigating the possibility of combining PCM and FDM systems in the same cable sheath. What these factors are, and the nature of their potential disturbing effect, may determine whether or not compatibility between the two systems is possible.

A unipolar pulse is shown in Figure 3. It represents a pulse such as it appears in the terminal equipment or in a regenerative repeater before the conversion to a bipolar format has taken place. It is an idealized pulse in that rise and fall times as well as aftershoot have been neglected. The duty cycle of the pulse train is 50 percent. This means that a unipolar string of binary ones in a terminal with a period T has a pulse width of T/2.

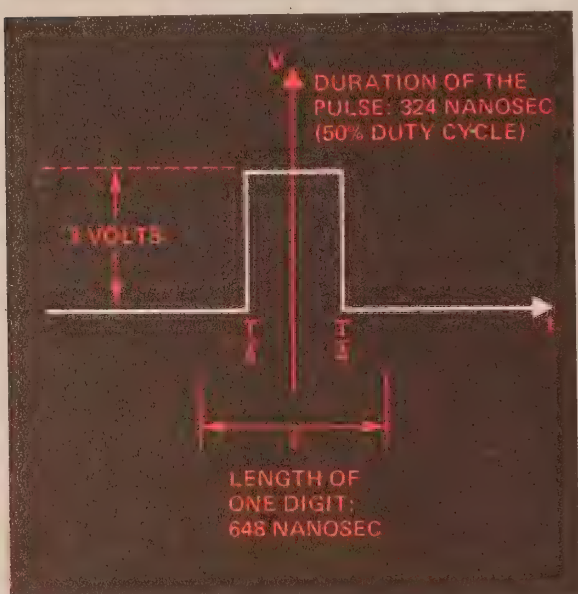


Figure 3. Unipolar pulse.

On a working D1/T1-type system with 24 voice channels carrying traffic, pulses occur randomly and with a pulse density closely approaching 0.50. That is, over a long enough period of time the number of binary ones and zeros (pulses and spaces) tend to be approximately equal. On a D2/T1-type system, pulses will tend to occur with a density somewhat greater than 0.50 (0.55 to 0.65 for busy-hour condition).

Before application to the transmission line, the pulse train is converted to a bipolar format by inverting every other pulse to opposite polarity (see Figure 4). The purpose of this inver-

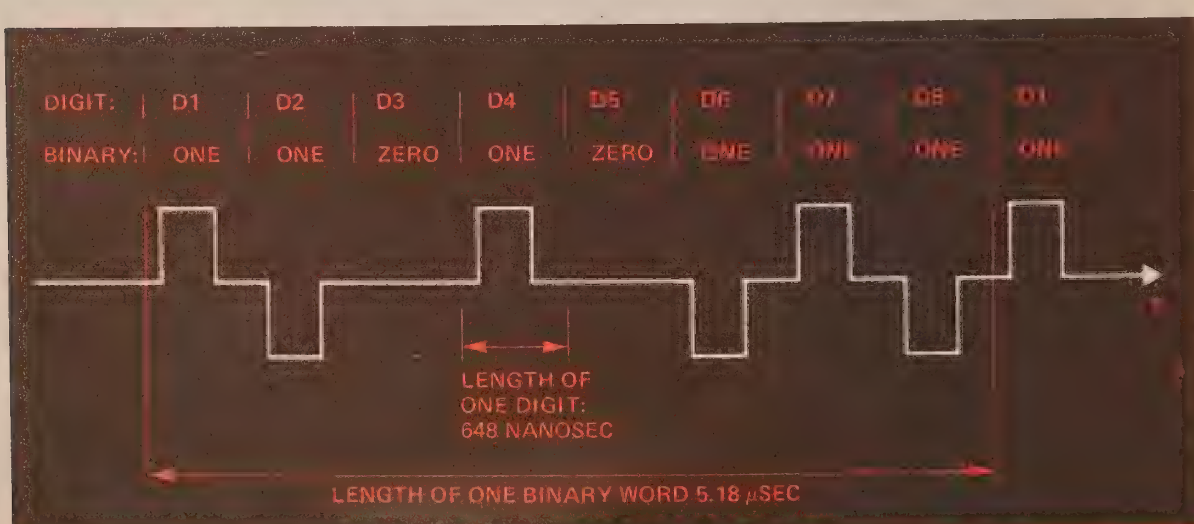


Figure 4. Bipolar pulse train.

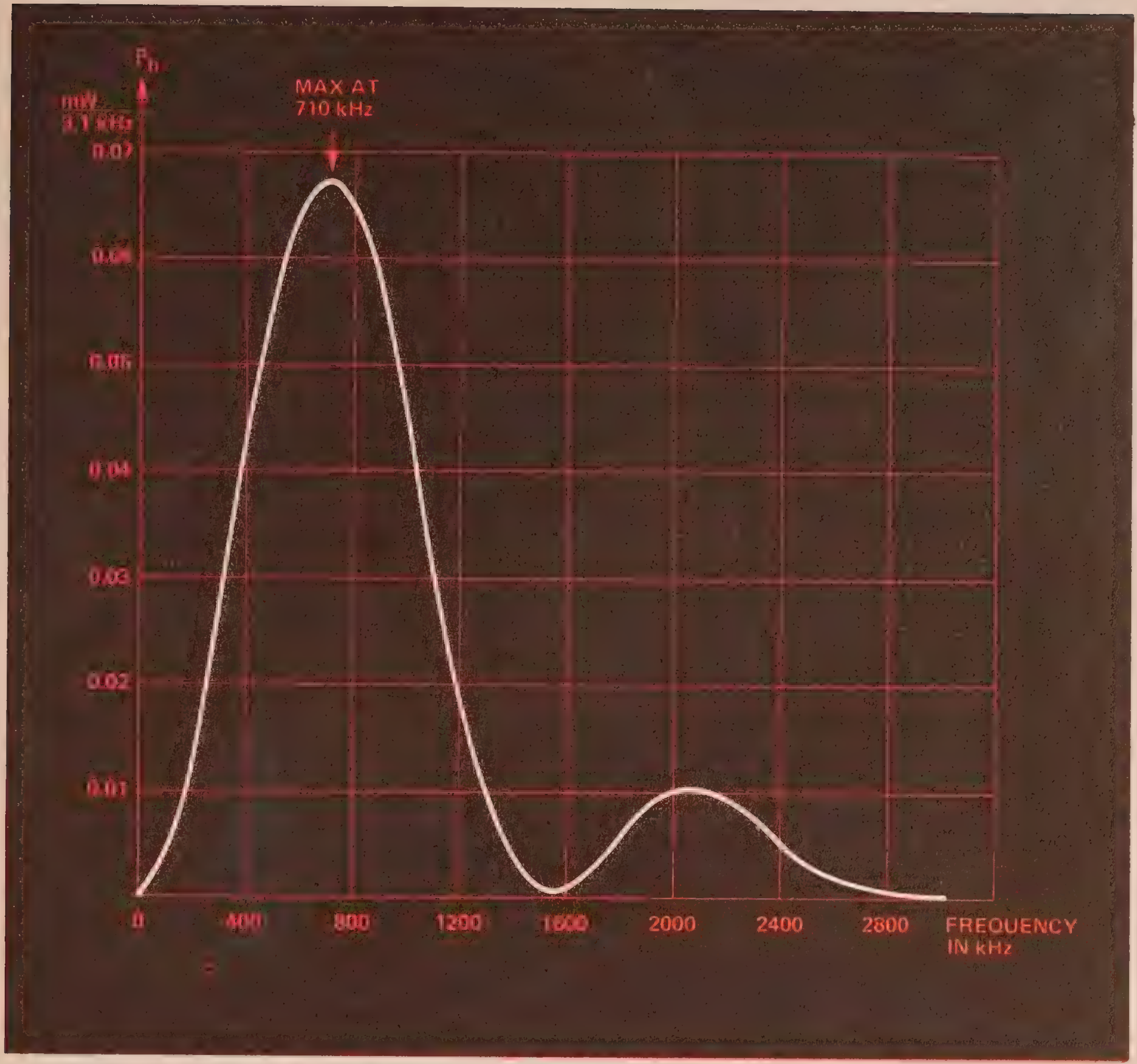


Figure 5. D1/T1-type power spectrum during busy-hour conditions.

sion is to shift the power spectrum to lower frequencies and to remove the dc component of the line signal.

The bandwidth usually available for transmission of voice information in most channels is 3.1 kHz. The random bipolar pulse train has its power distributed (in mW per 3.1 kHz slot) as a function of frequency as shown in Figure 5. This power spectrum is for a D1/T1-type PCM signal as it would appear during traffic conditions at the output of a regenerator. The curve represents the statistical average during traffic conditions and is not valid during idle or on-hook conditions. (The power spectrum during idle or on-hook conditions will be discussed in Part II.)

When PCM and FDM are combined within the same cable, the PCM power spectrum curve of Figure 5 will play an important part in evaluating crosstalk coupling between the two systems.

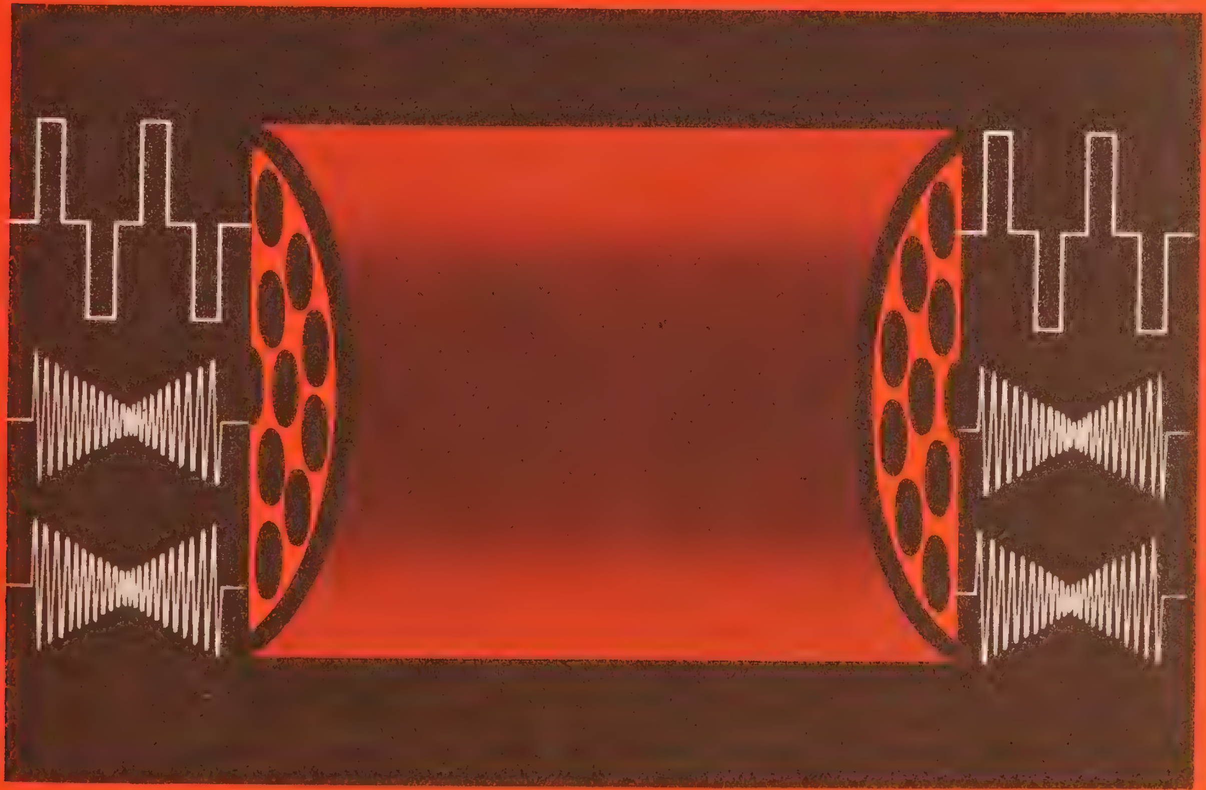
While this discussion of PCM-FDM compatibility has so far been of an empirical nature, Part II will investigate such factors as the PCM power spectrum and its effect of FDM systems, the effect of various types of PCM signaling on FDM, effects of idle/on-hook conditions and a method of estimating minimum crosstalk coupling loss.

GTE LENKURT

DEMODULATOR

AUGUST 1971

PCM-FDM Compatibility Part 2



The GTE Lenkurt study of PCM-FDM compatibility has yielded useful data in a quantity such that it will require three issues to cover all the information instead of the aforementioned two issues. Part II will deal with the theoretical aspects of PCM-FDM compatibility while Part III will apply these theories to a practical analysis of compatibility using a hypothetical PCM and FDM combination.

In the July issue (PCM-FDM COMPATIBILITY, PART I) of the Demodulator, the focus of attention was mainly on how to use to best advantage the empirical information thus far accumulated by GTE Lenkurt in the study of PCM-FDM compatibility. This included information on direction coordination of cable pairs within the same cable sheath and a discussion of the effects of far-end and near-end crosstalk on an FDM system.

This issue goes one step further in guiding the user toward achieving compatibility between PCM and FDM systems which lie within the same cable sheath.

Sampling, Quantizing and Encoding

The sampling, quantizing and encoding into digital form of an analog signal are the three major functions of a PCM terminal. The information used in this discussion is based on the sampling, quantizing, and encoding scheme used in the GTE Lenkurt 9001A, 9001B (both D1-type) and 9002A (D2-type) PCM channel bank assemblies. These assemblies are end-to end compatible with the Western Electric D1 and D2 channel bank assemblies and to similar terminals produced by other communications equipment manufacturers; hence the designation, "D1- and D2- type."

The level at which a voltage sample is quantized is relevant in evaluating PCM-FDM compatibility and is particularly important at the lower voltage levels since this is where the power spectrum may sometimes be confined to discrete frequencies during quiet and idle conditions in the PCM terminal. An "idle condition" implies that the telephone receiver may be on or off the hook and that no message is being transmitted even though there may be a line open between two parties.

In D1-type systems, each voice frequency channel is sampled 8000 times per second and each voltage sample has 127 discrete voltage levels available for quantization. The zero voltage level is known as level 64. There are 63 levels above level 64 in the positive direction, and 63 levels below level 64 in the negative direction. The number of the quantization level nearest the level sampled is encoded into a sequence of binary pulses and spaces, a pulse corresponding to a "one" and a space to a "zero". Figure 1A gives an example of noise or of low-level voice

signal sampling and also shows a portion of the level structure of a D1-type PCM system. Figure 1B shows examples of pulse patterns on the line. It should be noted that as the levels get further away from zero (level 64), the steps become progressively larger. This is because most of the information in speech signals is concentrated at low amplitude levels and small quantum steps are thus needed more at the low amplitude levels than at the higher levels in order to maintain a reasonably constant signal-to-noise ratio (in-

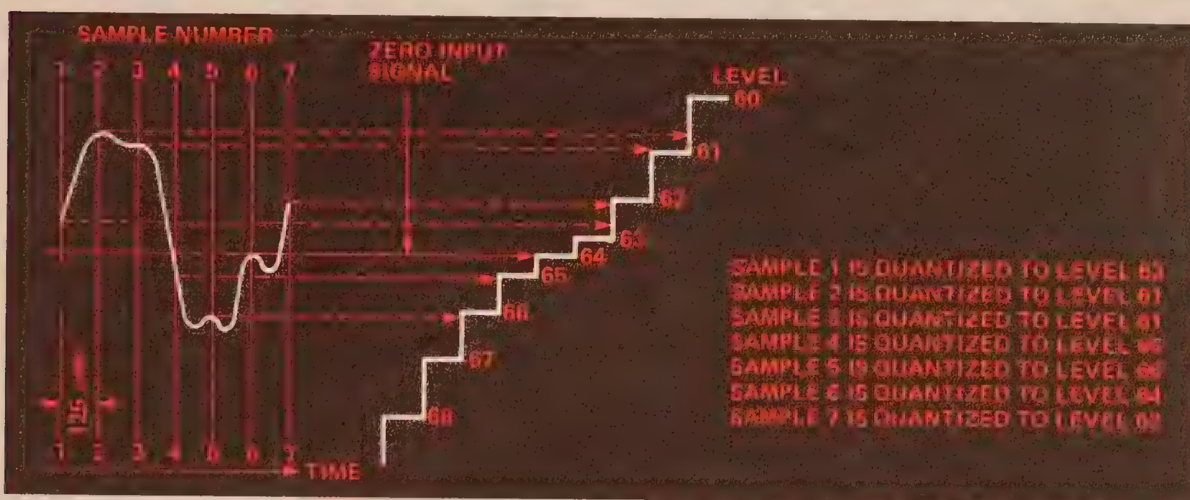


Figure 1A. In a D1-type terminal, each voltage sample is quantized (rounded off) to one of 127 possible levels. The number of the quanitzation level chosen is transmitted to the line as an encoded binary word.

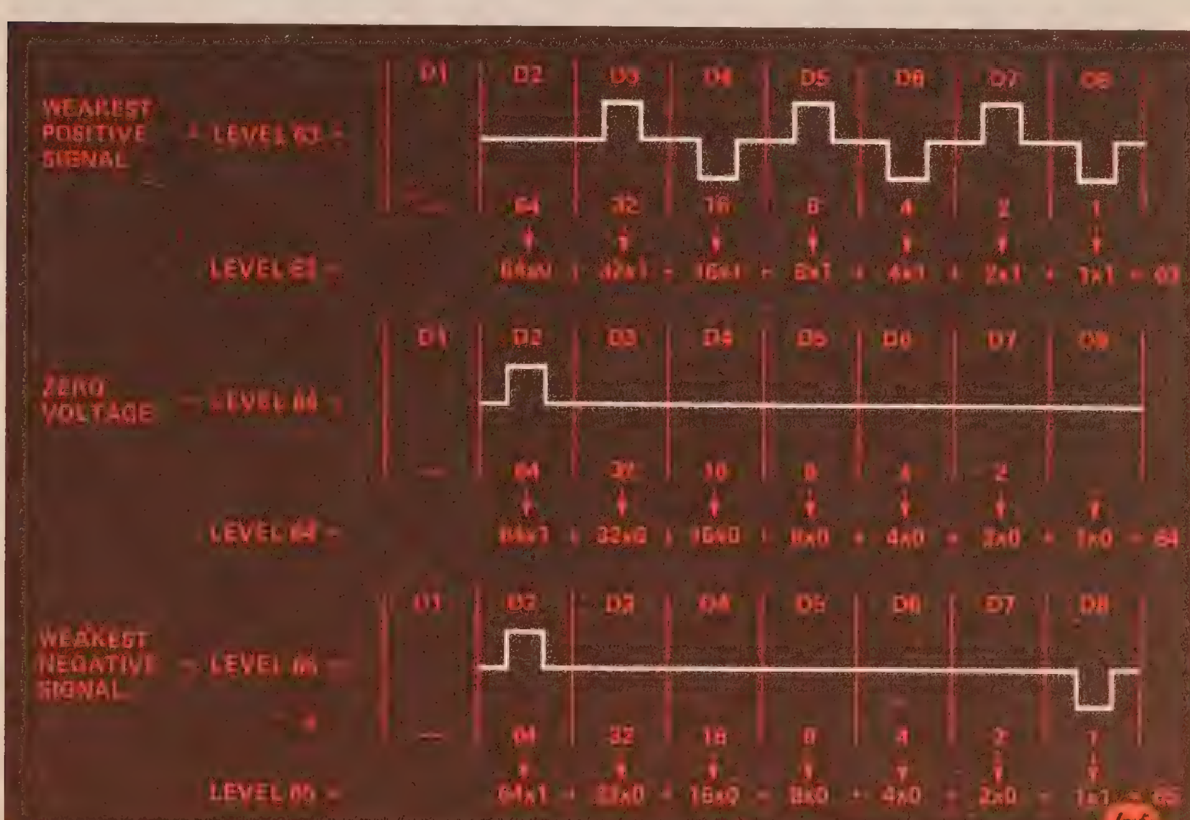


Figure 1B. Resulting pulse patterns on the line and corresponding calculations for small voltage samples in a D1-type terminal. (Positive and negative pulses both represent a binary one.)

dependent of signal level) which is the objective set for speech-loaded PCM telephone systems.

After the voltage is quantized (rounded off to the nearest quantization level), it is encoded into a 7-bit binary pulse pattern or binary word. A binary word consists of eight digits called D1 through D8. *The designation for the eight binary digits of the code word appear as D1 through D8 but should not be confused with the designation for D1- and D2- type terminals, they are two separate entities.* Digit D1 is used for signaling information only, while digits D2 through D8 represent the encoded version of the quantized sample (see Figure 1B). The 24 binary words representing the 24 voice channels plus the framing digit comprise one frame. There are 8000 frames per second and 8x24+1=193 bits per frame. Every other pulse is inverted to produce a bipolar pulse pattern for transmission.

In D2-type systems, the sampling rate is also 8000 times per second, but the number of quantization steps and the encoding method are different. In five frames out of every six all eight digits (D1 through D8) are used for encoding. There are then 255 discrete voltage levels available for quantization. These levels are numbered +0 to +127 and −0 to −127, zero quantization level corresponding to ±127. Figure 1C shows how a noise or low-level voice signal is sampled and quantized in a D2-type PCM system. Figure 1D shows examples of pulse patterns on the line. If the noise level in the terminal is sufficiently low to result in +127 or −127 quantized levels in every sample, a string of binary ones (pulses) broken by a zero (space) on the average only once every sixteen digits will be produced. In the sixth frame, one digit is used for signaling information so that only seven digits (127 levels) are available for quantizing. The availability of eight digits for quantization 5/6ths of the time provides for better signal quality than in a D1-type system.

Power Spectrum Curves

The power spectrum curve is an important tool in evaluating PCM-FDM crosstalk since from it can be derived the amount of potential interference to an FDM system channel which may be transmitted at a certain frequency.

The power spectrum (power as a function of frequency) of a D1/T1-type PCM bipolar pulse train as it would appear during traffic conditions at the output of a regenerator is shown in Figure 2. The curve is calibrated in dBm per 3.1-kHz slot. For this discussion, only a portion of Figure 2 is necessary since the maximum disturbance generated into an FDM line by a PCM system will occur at approximately 710 kHz and the FDM cable carrier systems of most concern for this discussion occupy the frequency range under 400 kHz. Figure 3 shows the significant portion of the power spectrum curve in expanded form. Also shown in expanded form, is the curve for a D2/T1-type system. Although the D1- and D2-type terminals cannot be operated end-to-end, they can both be operated over the T1-type repeatered line.

The power spectrum for a D1/T1-type system is based on a pulse density value of p=0.50, where p is the probability of a binary one (a pulse). This value for p is quite constant with varying loads in a D1/T1-type system provided that there is traffic on at least six of the 24 channels in the terminal. The p=0.50 implies that there is an equal probability of a one or a zero in the pulse train.

The curve for a D2/T1-type power spectrum is based on a pulse density of p=0.55 during busy-hour traffic condi-

tions. When loading decreases, p (pulse density) will increase, resulting in a decrease in P_b (power) at low frequencies (below approximately 420 kHz) and an increase around 710 kHz.

D1A and D1B Signaling

In the study of PCM to FDM interference, the type of signaling used in the PCM system will dictate the fundamental frequency or multiple thereof at which the greatest interference will occur when the PCM system is in the idle condition (no traffic).

Two types of signaling are used with D1-type terminals. These two signaling configurations bear the Western Electric designation of D1A and D1B.

When using D1A (also called "D1/D8") signaling the D1 digit in every binary word is used once in every frame to convey the signaling information. In the on-hook condition there is, then, always a pulse (a binary one) in the D1 time slot in every frame for a given channel. A problem arises when for certain signaling requirements (foreign exchange or revertive pulse signaling, for example) a second signaling channel is required. If this happens during voice transmission

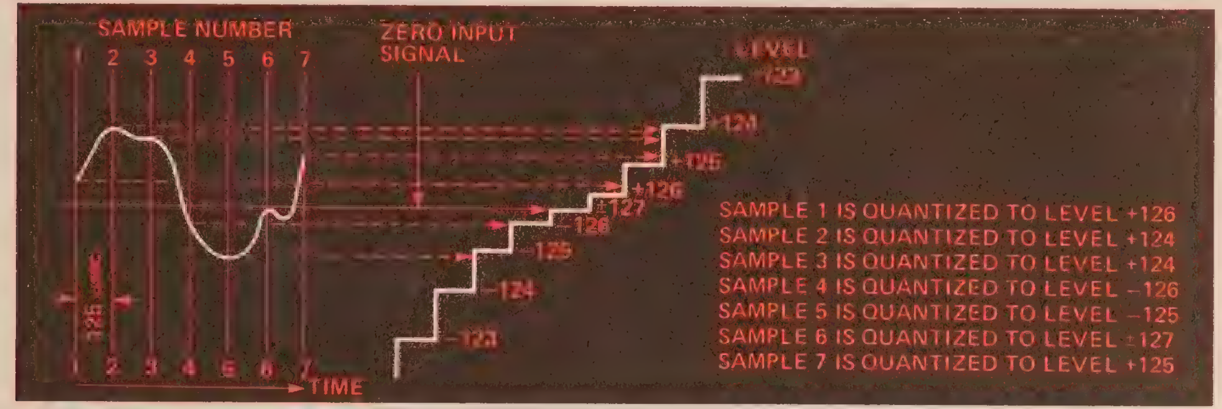


Figure 1C. In a D2-type terminal, each voltage is quantized to one of 255 possible levels. -

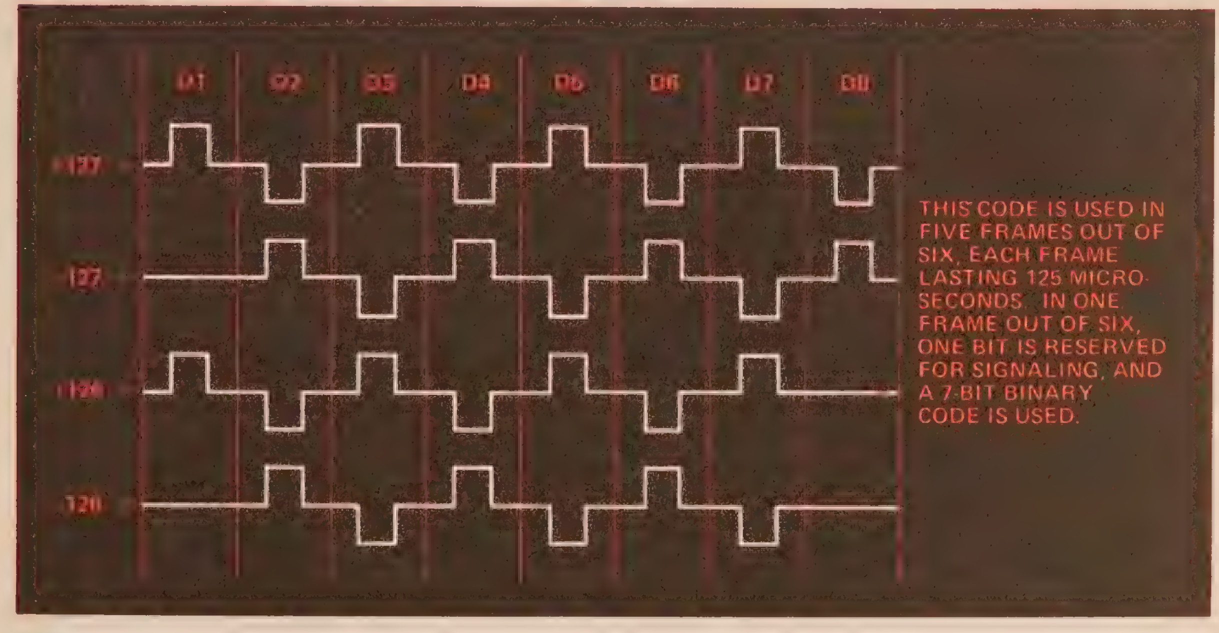


Figure 1D. Pulse patterns resulting from small voltage samples in a D2-type terminal will result in a pulse density close to 100 percent.

(this can happen, for example, when there is no answer supervision), only digits D2 through D7 (six digits) are available for representing the quantized sample. Six-digit encoding corresponds to 63 levels as compared to the usual 127 levels for seven-bit encoding. This six-digit encoding results in larger steps between levels and consequently in greatly increased quantizing noise within the PCM system itself.

To avoid any increase in noise due to requirements for more than one signaling channel, the D1B (also called "D1 only") signaling arrangement was developed. With D1B signaling, the signaling rate is divided by a factor of four so that the D1 digit is used for signaling information only once every fourth frame (per signaling channel). This effectively creates the potential

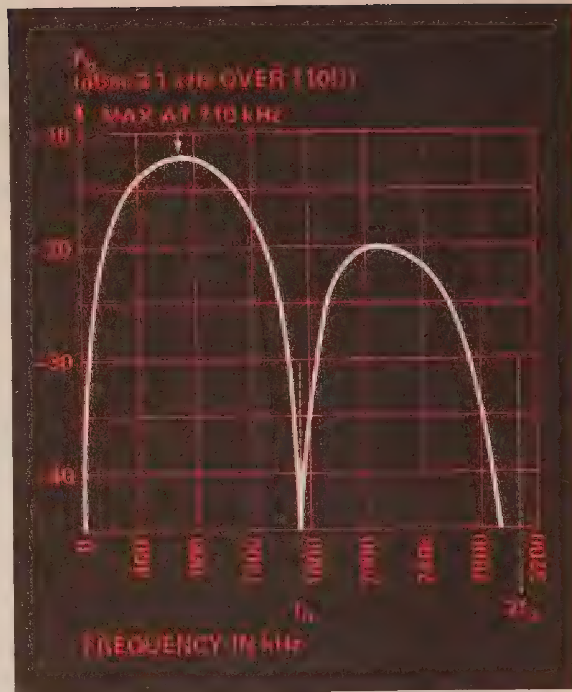


Figure 2. Power spectrum of a D1/T1 type system at the output of a regenerator during traffic conditions.

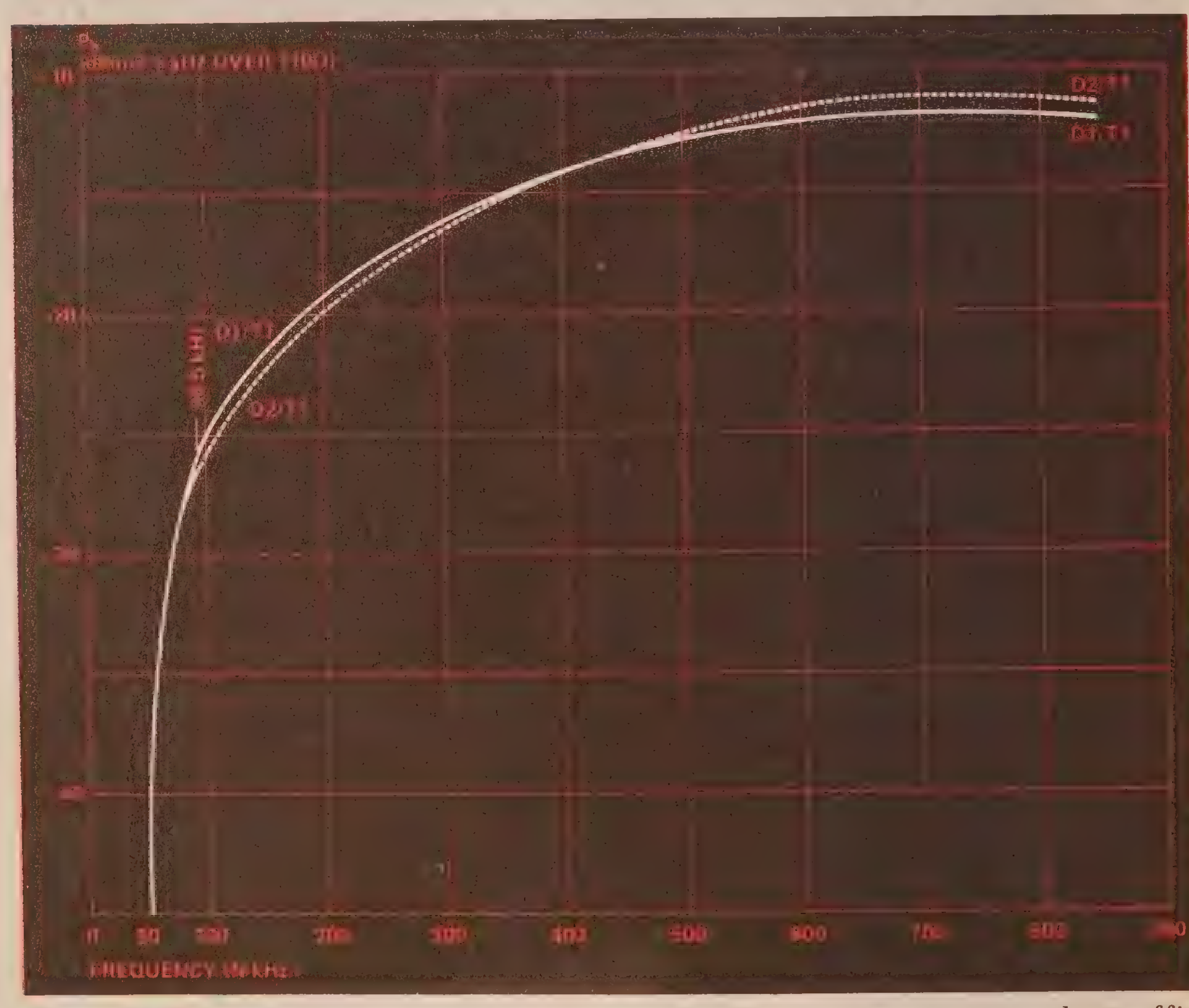


Figure 3. Expanded power spectrum curve of a T1-type pulse train under traffic conditions.

for four signaling channels instead of one, although as a rule, not more than two channels are used. (See the March, 1971, issue of the GTE Lenkurt Demodulator for an extensive treatment of PCM signaling.)

Cable Characteristics

Aside from removing the dc component from the line, conversion of the PCM pulse train to a bipolar format also shifts the power spectrum to lower frequencies. This shift in power spectrum is advantageous for PCM systems because the crosstalk characteristics of cables are better at lower frequencies. Operation at lower frequencies also results in relaxed requirements on cable makeup, pair selection, and/or repeater spacing.

While the impedance of a cable pair is mainly a function of frequency, it also depends on such factors as cable gauge, insulation, and capacitance. Cable impedance falls rapidly from a value of 600-900 ohms at voice frequencies to approximately 100 ohms at 300 kHz and stays relatively constant above that frequency. A compromise value of 110 ohms has been chosen for the 50-400 kHz region which is the band of greatest interest for this study. This compromise impedance is accurate within this frequency range to within approximately 10 percent.

Idle/On-Hook Pulse Pattern

The idle/on-hook pulse pattern is important in the study of interference since an idle and quiet PCM system (all 24 channels idle) can sometimes (if the noise level is sufficiently low) produce a repetitive pulse pattern (resulting in a power spectrum confined to discrete frequencies only) instead of the random distribution of binary ones and zeros normally present when computing the power spectrum. This can cause excessive interference at certain discrete frequencies. A repetitious pulse pattern will appear on digits D2 through D8 in a D1-type system if a very low noise level is present at the terminal. *The value of the D1 digit in the idle condition is determined for the D1-type terminal only by the on-hook or off-hook condition of the channels.*

D1A Idle/On-Hook

If a D1-type system is arranged for D1A signaling, and a condition exists in which all channels are on-hook, with a very low noise level in the PCM terminal, there will be pulses for all D1 and D2 digits in every frame (digits D3 through D8 being zeros or spaces) since this sequence for digits D2 through D8 represents a zero input level (level 64) and since D1 is used strictly for signaling information (the on-hook condition is represented by a binary one). Such a repetitive pulse pattern has a line power spectrum (power concentrated at discrete frequencies). This means that the power in the signal can be represented by components of power at discrete frequencies which are multiples (harmonics) of a fundamental frequency, in this case, 193 kHz. This component may create serious interference in the 36-268 kHz band occupied by an N-type FDM carrier system, for example.

The fundamental frequency is derived for a periodic pulse pattern by the formula:

$$f_o = \frac{1}{T} = \frac{1}{5.18 \times 10^{-6}} = 193 \text{ kHz},$$

where T is the length of one period in seconds.

The periodic idle/on-hook pulse pattern and its corresponding line power spectrum for D1A signaling appear as shown in Figures 4A and 4B. A PCM terminal in a telephone office generally picks up some noise from

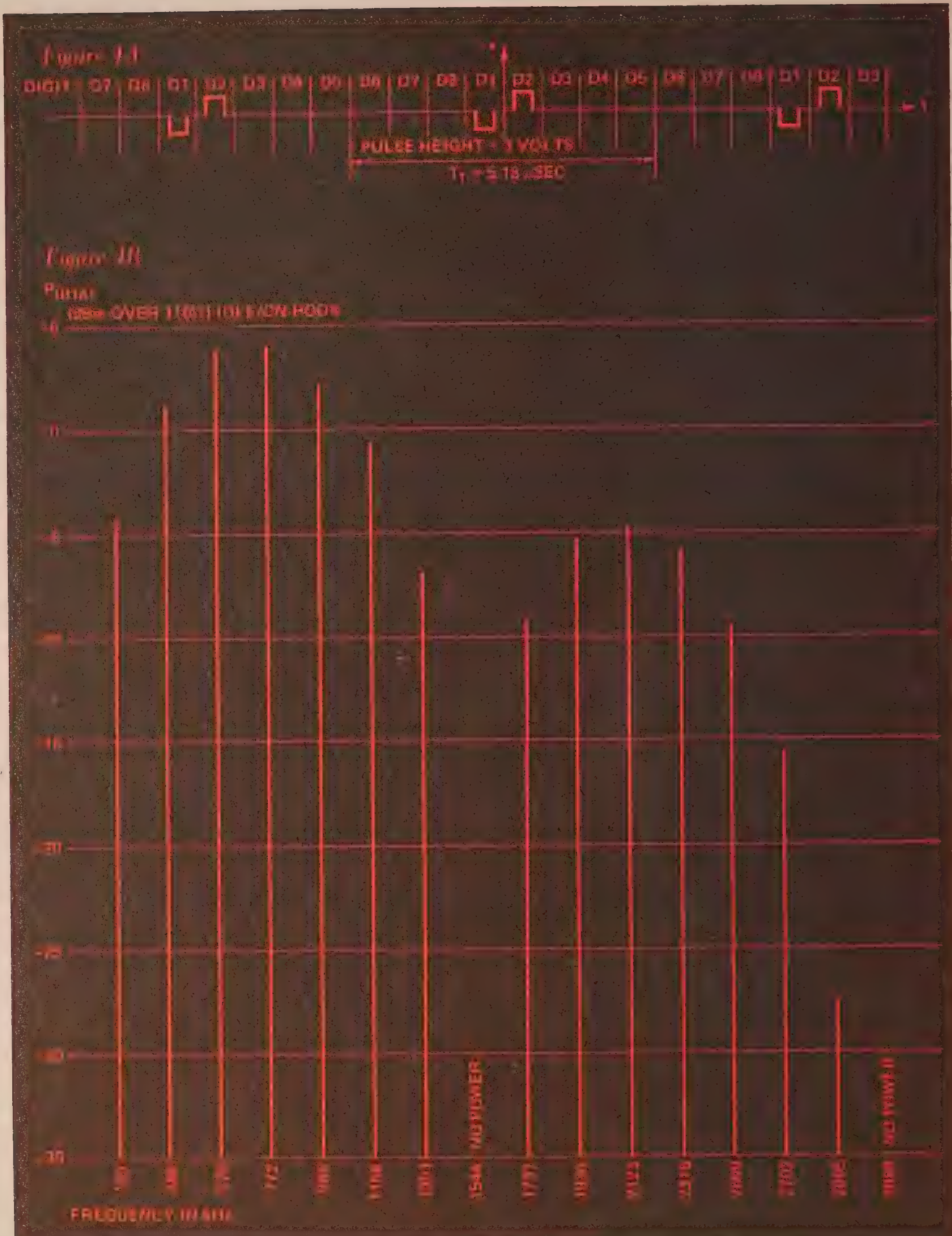
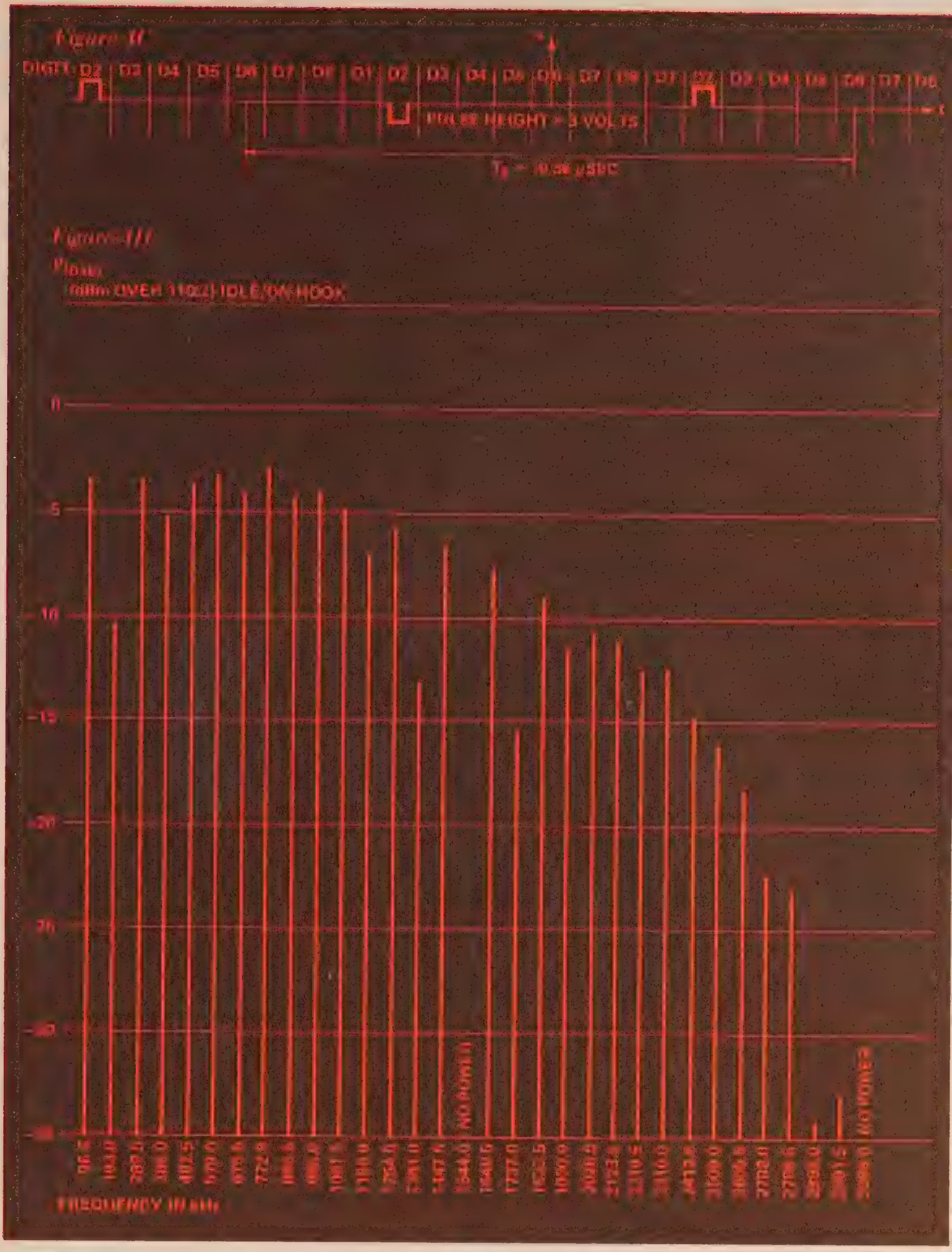


Figure 4A shows the pulse pattern as it appears out of a PCM terminal or regenerator in the idle/on-hook condition when D1A (D1/D8) signaling is used. The resulting line power spectrum is shown in Figure 4B.

Figure 4C represents the pulse pattern in three frames out of four for the idle/on hook condition and D1B signaling.

Figure 4D is the line power spectrum resulting from the idle/on-hook pulse pattern in a D1-type terminal when D1B (D1 only) signaling is used. That pattern is according to Figure 4A every fourth frame; according to Figure 4C the remaining three (out of four) frames.



switching transients, which causes the power spectrum on the transmission line to be less clean-cut than that shown in Figure 4B. In most working systems, noise will cause the quantized samples of the input signal to fluctuate randomly around level 64. These random fluctuations will introduce more pulses per binary word and thus more high-frequency components to the line. Even rather small noise levels (for example, resulting in levels 63 or 65 most of the time, rather than 64) will result in a pulse density of approximately 50 percent. *Only when all voltage samples are consistently quantized to zero level does the power spectrum become a series of spectral lines.*

Figure 5. Power spectrum for a working PCM system during a busy-hour condition (10 a.m.) on a week day.

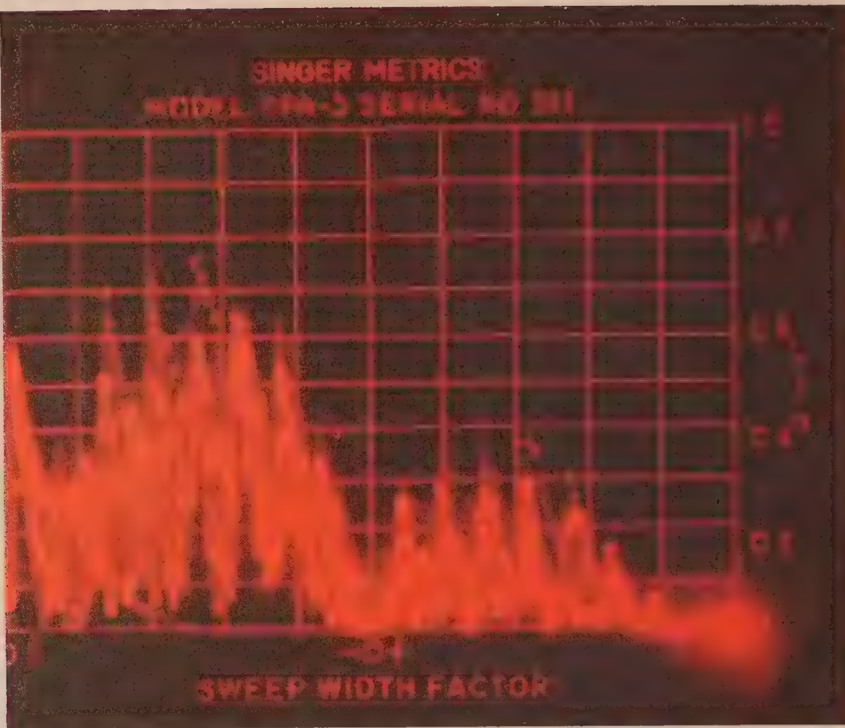


Figure 6. Power spectrum for a working PCM system during a low traffic condition (midnight).

Figures 5 and 6 show the power spectrum of the line signal as it appears at the output of a PCM terminal under different traffic conditions. These photographs were taken in the field on a working 24-channel D1A system. Figure 6 shows that the power spectrum only *approaches* a spectral line condition due to the presence of ambient noise. However, a spectral line condition can be attained in the laboratory where ambient noise is more strictly controlled.

Although a periodic pulse pattern is a rather unusual case since it corresponds to all channels off hook and a very low noise level, it is important because it usually represents the worst

interference condition as far as FDM channels in the corresponding discrete frequency slots are concerned. As the pulses occur randomly with traffic, the power spectrum of a PCM system will tend to assume a smooth curvature as shown in Figure 3. Figure 4B represents the worst case for the all channels on-hook condition, as far as crosstalk at low frequencies is concerned.

D1B Idle/On-Hook

When a terminal is arranged for D1B ("D1 only") signaling, the idle/on-hook pulse pattern is somewhat more complicated than for D1A signaling. In one frame out of four the pulse pattern will be as shown in Figure 4A; in three frames out of four it will be as shown in figure 4C. The pulse train in one frame out of four for the idle/on-hook condition is thus identical to that for D1A signaling. This will produce lines in the power spectrum at multiples of 193 kHz as for D1A signaling, but their magnitudes are reduced by a factor of four (6dB) because it represents only the pulse pattern in every fourth frame. The important thing to be noted about the power spectrum for an all channels idle and on-hook condition, (as in Figure 4D) is that it consists of two sets of spectral lines for D1B signaling. One set has lines at all odd multiples of 96.5 kHz caused by the pulse pattern in three out of four frames. The other set has lines at even multiples of 96.5 kHz (which is the same as multiples of 193 kHz) caused by the pulse pattern in one out of every four frames.

D2/T1-Type Idle/On-Hook

The pulse pattern for the idle/on-hook conditions for a D2/T1-type system is a sequence of consecutive pulses (a string of binary ones), broken by an occasional space. If the system is idle with all channels in the on-hook condition, and the noise level is sufficiently low, the voltage samples will all be quantized to +127 or −127 in five frames out of six; to +63 or −63 in one frame out of six. The signaling digit for the on-hook condition is a binary one. For this condition, on the average 15 out of every 16 digits will be binary ones. Analysis of this type of pulse train shows a reduction of single-tone interference below 400 kHz amounting to 11 dB (compared to idle/on-hook D1A) or 7 dB (compared to idle/on-hook D1B).

Crosstalk Evaluation

From a study of idle/on-hook conditions for D1A and D1B signaling it is apparent that PCM to FDM interference will show greatest potential for disturbance at multiples of 96.5 kHz or 193 kHz during the times of the day or night when most of the PCM channels are idle. The value of the worst case component under this condition can be obtained from the applicable diagram in Figure 4. It is possible that this disturbing effect from the PCM system in its idle condition will in some cases necessitate the elimination of FDM channels in baseband slots at or adjacent to multiples of 96.5 kHz.

In Part III of PCM-FDM COMPATIBILITY, the theoretical aspects of compatibility which have been discussed in the previous two issues of the Demodulator will be put to practical application. A hypothetical PCM-FDM combination will be evaluated to determine if the two systems are compatible.

GTE LENKURT

DEMODULATOR

SEPTEMBER 1971

PCM-FDM Compatibility Part 3

Subjection of an FDM system to interference from PCM signals is the major factor which discourages the user of communications equipment from placing PCM and FDM systems in the same cable sheath. By giving careful consideration to direction coordination of cable, and by evaluating the effects of PCM to FDM crosstalk, single-tone interference, and length of repeater sections, the user may find his FDM system compatible with PCM without extensive modifications.

Parts I and II of the PCM-FDM compatibility series dealt with the empirical and theoretical aspects of PCM and FDM systems which lie in the same cable sheath. Some of the factors influencing the amount of interference between these two types of systems are crosstalk coupling loss, length of exposure of FDM to PCM, the frequency band occupied by the FDM system, and the type of PCM signaling utilized. The discussion of these subjects brings to light a way of evaluating the amount of interference which the FDM system receives from the PCM system.

A series of steps may be taken to evaluate the possibility of compatibility between PCM and FDM systems which operate on pairs within the same cable sheath. To demonstrate how this series of steps can be useful to the communications equipment user, an example is given in this issue (using a hypothetical PCM-FDM combination) on the process of estimating the minimum crosstalk coupling loss required if two systems are to be compatible.

The example for crosstalk evaluation will employ a GTE Lenkurt 24-channel 91A PCM system (D1/T1-type) equipped for D1A signaling and a GTE Lenkurt 47A (an N-type FDM system) equipped with compandors. The calculations are valid for any D1/T1-type PCM system and any N1 or N2-type FDM system.

The 47A is a 12-channel, double-sideband, amplitude-modulated carrier system which operates over two cable pairs, one for each direction of transmission. It is end-to-end compatible with Western Electric N1 or N2 systems (depending on the option of 47A used). Compandors are optional for these systems. On any section of the cable, the two directions of transmission utilize different frequency line groups. The low-frequency group extends from 40 to 128 kHz and the high-frequency group from 176 to 264 kHz as shown in Figure 1. In each repeater there is a modulator which shifts the signal from one band to the other (frequency frogging) in order to combat near-end crosstalk.

For this example it is assumed that maximum-length repeater section lengths are used for the 47A system (40 dB at 176 kHz) and that end sections (or any sections adjacent to a telephone switching office) do not exceed 25 dB at 176 kHz. The 40-dB value is the maximum allowable power loss over an intermediate repeater section where little interference will be encountered from external sources; for this discussion an intermediate repeater location not at a switching center will be referred to as a "low-noise point." The 25-dB value is the maximum allowable power loss over a repeater section adjacent to an office (an end section, for example). At an office the repeater is subject to interference from office switching equipment; for this discussion offices will be referred to as "high noise points." For 22-gauge PIC (polyethylene insulated

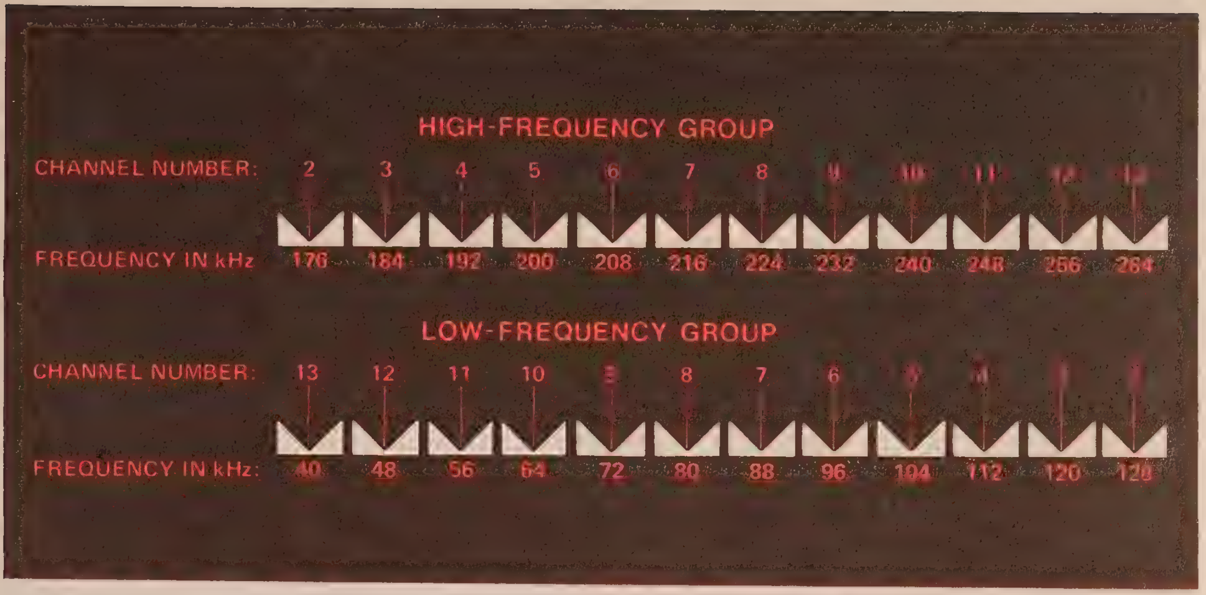


Figure 1. Frequency designations for the GTE Lenkurt 47A FDM System. Channel 1 (not shown) is an optional channel which is seldom used due to performance limitations at that frequency although it may be used in lieu of any of the other available channels.

cable) with a capacitance of 0.083 μF per mile, the 40- and 25-dB values correspond to repeater section lengths of 3.88 miles and 2.42 miles, respectively. The loss at 772 kHz (the loss at this frequency determines the PCM repeater spacing) is 90 dB over a 3.88-mile section of such cable. Assuming that the PCM intermediate repeater sections are exactly one third of the 47A repeater section (this corresponds to 30 dB at 772 kHz), that the 47A repeater locations within the section exposed to PCM interference always coincide with a PCM repeater point, and that the exposure to PCM occurs over three 47A intermediate repeater sections plus one end section, the system layout will appear as in Figure 2.

Noise Interference

The interference from a PCM system consists either of noise or single-tone interference depending on the traffic loading of the PCM system and the noise level at each PCM terminal. The first consideration will be that of noise interference from a traffic-loaded PCM system.

The near-end crosstalk interference from a PCM system to an FDM system is shown in Figure 2 by the green arrows which indicate the near-end crosstalk paths of importance. The interference from PCM is most severe at the high-level outputs of the PCM repeaters (regenerators). However, the only near-end crosstalk of importance occurs on the PCM repeater sections adjacent to FDM repeater receive-inputs, as the green arrows in Figure 2 indicate. This is because the level of the received signal on the FDM system is the lowest and most noise-sensitive at that point. The PCM to FDM near-end crosstalk originating on PCM repeater sections not adjacent to an FDM repeater receive-input will be attenuated by 10 dB or more before reaching the FDM repeater input and can be neglected.

The allowable degradation of noise performance in the 47A system has been chosen such that the presence of PCM carrier interference should not cause the noise performance to deteriorate to worse than 27 dBrnc. This is one dB worse than the worst line-up

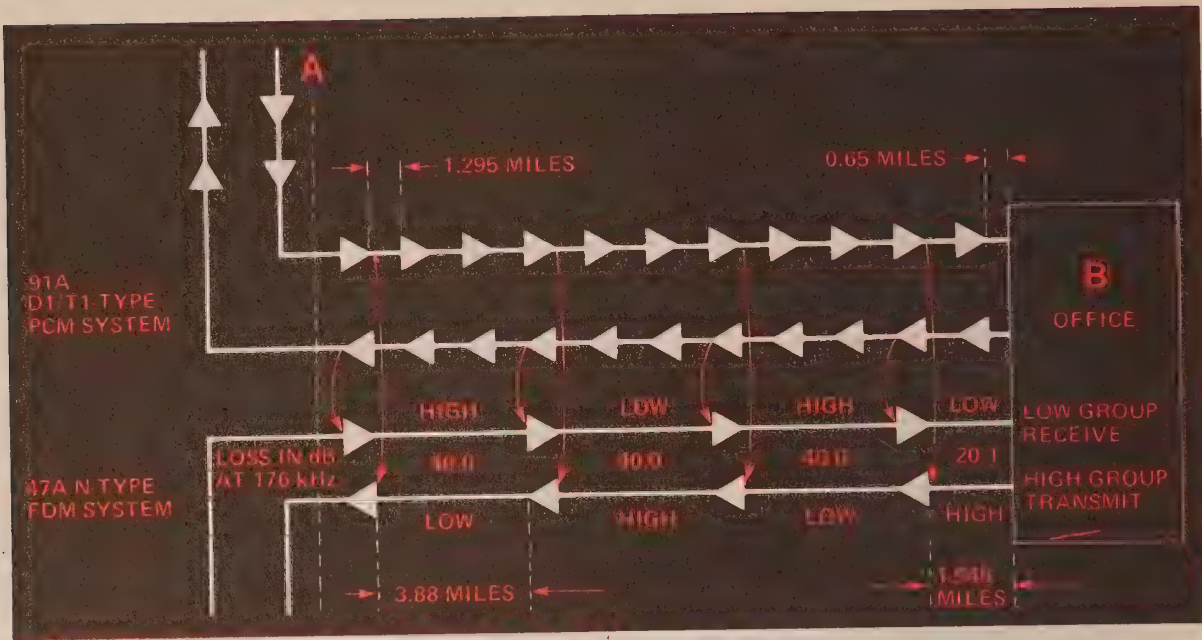


Figure 2. PCM-FDM showing three intermediate repeater sections and one end section of the FDM system exposed to PCM interference.

noise performance allowed for a long distance 47A system without PCM interference (26 dBrnc). Based on this requirement, the total noise contribution originating from PCM carrier systems must not exceed 20 dBrnc, since by dB addition laws, 26 dBrnc + 20 dBrnc = 27 dBrnc. The dBrnc unit is used to measure absolute noise and from the conversion table shown in Figure 3, a 20-dBrnc value converts to 100 pW psophometrically weighted (100 pWp).

A performance requirement of 20 dBrnc for PCM interference corresponds to an FDM signal-to-PCM interference noise ratio (test tone-to-PCM interference noise ratio) of 68 dB (as shown in Figure 3). The expression "signal-to-noise ratio" generally used in FDM system terminology actually means test tone-to-noise ratio since it refers to a test tone which is injected into the system from a signal generator for measurement purposes. At test tone level, a 47A non-compressed carrier is amplitude-modulated with a modulation index of 0.35 (35 percent). This 35% value was chosen in the initial system design to avoid exceeding 100 percent modulation at even the highest speech volumes.

Figure 4 shows the relationship between carrier and test tone signals in one channel of an FDM system. In order to solve for unknown noise levels a meaningful relationship must be established between the carrier-to-noise and signal-to-noise ratios. For this non-compressed, double-sideband, amplitude-modulated, 47A FDM system, in which a test tone modulates the carrier 35 percent, the following equation is true:

Equation 1

$$\left(\frac{C}{N}\right)_{6.2\ \text{kHz}} = \left(\frac{S}{N}\right)_{3.1\ \text{kHz}} + 9\ \text{dB}$$

where C/N stands for FDM carrier-to-PCM interference noise ratio (where the carrier is at one specific frequency and the PCM noise is over 6.2 kHz) and S/N for FDM test tone-to-PCM interference noise ratio, with reference to the voice frequency drop point. The 3.1- kHz value, corresponds to the usable sideband bandwidths in a voice channel. Since the 47A uses double

dBrnc	pWp	S/N (dB)
6	4.0	82
7	5.0	81
8	6.3	80
9	8.0	79
10	10.0	78
11	12.6	77
12	15.9	76
13	20.0	75
14	25.1	74
15	31.0	73
16	39.8	72
17	50.0	71
18	63.0	70
19	79.5	69
20	100.0	68
21	126.0	67
22	159.0	66
23	200.0	65
24	251.0	64
25	316.0	63
26	398.0	62
27	500.0	61
28	630.0	60
29	795.0	59
30	1000.0	58
31	1260.0	57

All units are given with reference to test tone level of zero dBm.

dBrnc — C-message weighted.

pWp — Picowatts psophometrically weighted.

Figure 3. Noise measurement conversion table.

sideband operation, each carrier is associated with noise interference over a bandwidth of 6.2 kHz. Equation 1 states that carrier-to-noise is equal to the signal-to-noise + 9 dB.

The maximum allowable noise performance for a GTE Lenkurt 47A system is 26 dBrnc. In a case where the repeater sections are of acceptable lengths and good cable and line-up procedures are used, the noise performance will usually be much better than 26 dBrnc. However, this discussion will proceed as if the noise performance on the system was at the 26-dBrnc point (worst case) before the addition of PCM interference. In order to allow for this PCM interference, one additional dB of interference will be accepted which will make the total noise performance 27 dBrnc. For this discussion the 20-dBrnc value will be regarded as the total noise interference to the FDM system although it must be remembered that a 26-dBrnc noise value *does* exist in addition to the 20 dBrnc.

The 20-dBrnc PCM noise requirement previously calculated corre-

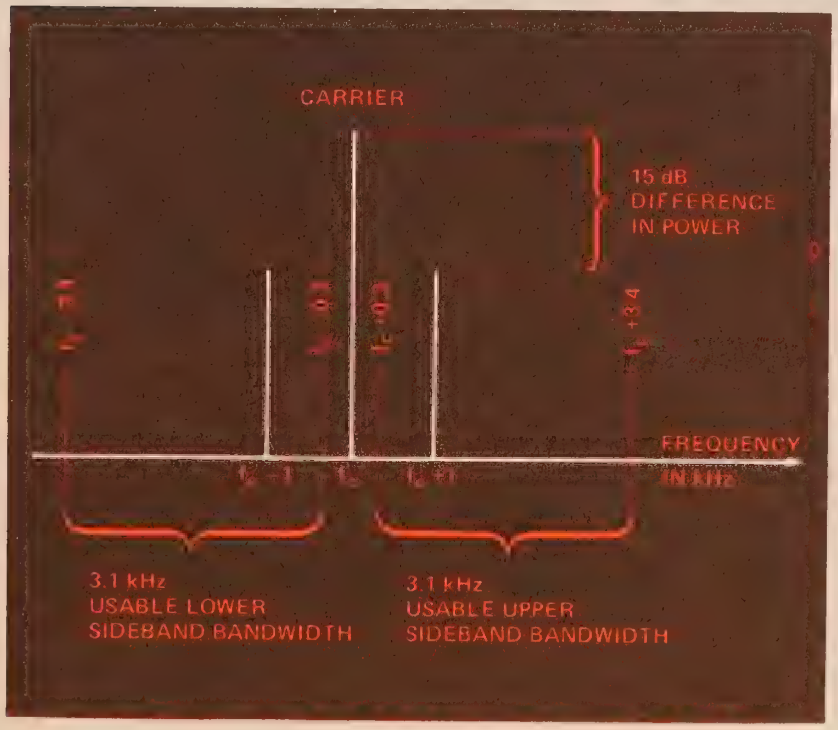


Figure 4. When a carrier of frequency f_c is modulated by a sinusodial 1-kHz test tone, sideband frequencies appear at frequencies (f_c-1) kHz and (f_c+1) kHz. Each one of these two sideband frequency components are of a power level 15 dB below the level of the carrier. This is based on the test tone modulating the carrier with a modulation index of 0.35 (such as in a non-compandored 47A system).

sponds to an FDM signal-to-PCM interference noise ratio of 68 dB, and thus converts to a requirement for FDM carrier-to-PCM interference noise ratio of:

$$\left(\frac{C}{N}\right)_{6.2\ \text{kHz}} = \left(\frac{S}{N}\right)_{3.1\ \text{kHz}} + 9\ \text{dB}$$

$$\left(\frac{C}{N}\right)_{6.2\ \text{kHz}} = 68 + 9\ \text{dB}$$

$$\left(\frac{C}{N}\right)_{6.2\ \text{kHz}} = 77\ \text{dB; (non-compressed).}$$

For this discussion a compandor will be inserted in accordance with the original conditions at the beginning of this example. This produces a 20-dB compandor advantage and results in a requirement of:

$$\left(\frac{C}{N}\right)_{6.2\ \text{kHz}} = 57\ \text{dB (compressed).}$$

The carrier with the lowest power level will differ in level and frequency depending upon the 47A repeater section length and whether the low-group or the high-group is being observed. Levels will thus be different on repeater sections adjacent to switching centers (as in end sections) compared to intermediate sections adjacent to low-noise points. If the spacing rules for N-type carrier are adhered to (end-sections not to exceed 25 dB at 176 kHz, intermediate sections not exceeding 40 dB), the lowest carrier levels on a 47A system have been calculated to be approximately:

(a) At a high-group repeater input at the high-frequency end of the band (264 kHz) . . . −48 dBm

(b) At a low-group repeater input at the high-frequency end of the band (128 kHz) . . . −42 dBm

(c) At a high- or low-group repeater input or terminal input located in or adjacent to a switching center . . . . . . −28 dBm.

These "lowest" FDM carrier power levels are significant when evaluating PCM to FDM interference since it is at these levels that an interfering PCM system will have the greatest effect on FDM. It is assumed throughout the following calculations that the spacing rules quoted above were followed when laying out the N-type carrier system.

A repeater input on a section adjacent to a switching center is at least 14 dB less sensitive to PCM interference than other intermediate sections due to its shorter length (compare values −28 dBm versus −42 dBm and −28 dBm versus −48 dBm). End-sections and other sections adjacent to high-noise points can thus be neglected for the purpose of these interference calculations if there are at least as many normal (adjacent to low-noise points) intermediate 47A repeater sections in the section being exposed to PCM interference as there are 47A repeater sections adjacent to high-noise points. The approximation error thus incurred is no more than 0.2 dB.

If a 47A carrier system is exposed to PCM interference over several intermediate sections (as in this example), about half the number of sections exposed (for any direction of transmission) are high-group sections and half are low-group sections. Consideration must therefore be given to the relative effects of interference into these two groups.

High-Group Repeater Input

The minimum FDM carrier-to-PCM noise ratio requirement has been calculated to be 57 dB (68 dB + 9 dB − 20 dB = 57 dB). The lowest power level (this is the worst acceptable case con-

dition) of the highest-frequency carrier (at 264 kHz) is –48 dBm as previously mentioned. The maximum allowable noise level due to PCM carrier interference in the 47A FDM system is then calculated as follows:

$$\left(\frac{C}{N_{max}}\right)_{6.2\ kHz} = 57\ dB.$$

By virtue of the fact that C is –48 dBm in the worst case, and that the dB division rules state that division of two values effectively means subtraction of these values when expressed in dB form, it follows that,

$$-48\ dBm - \left(N_{max}\right)_{6.2\ kHz} = 57\ dB$$

$$\left(N_{max}\right)_{6.2\ kHz} = -48\ dB - 57\ dB$$

$$= -105\ dBm$$

or

$$\left(\bar{N}_{max}\right)_{3.1\ kHz} = -108\ dBm$$

where N_{max} is the maximum allowable noise level in the 6.2-kHz or 3.1-kHz slots at 264 kHz. Dividing the bandwidth by 2 makes the requirement more severe by 3 dB and hence, N_{max} in a 3.1 kHz slot becomes –108 dBm.

The noise power in a 3.1 kHz slot centered at 264 kHz of the PCM system is –17 dBm according to the power spectrum curve of Figure 5. The difference between –17 dBm and –108 dBm is 91 dB. This would be the requirement for near-end crosstalk coupling loss at 264 kHz between PCM and FDM cable pairs in this example if the following were true:

(a) The effect of far-end PCM to FDM crosstalk could be neglected

(b) Only one PCM system interfered with the FDM system

(c) Only one 47A repeater section were exposed to PCM interference.

Also, the effects of single-tone interference have been neglected up to this point but will be covered in following paragraphs along with the effects of a PCM system interfering with more than one FDM repeater section.

Low-Group Repeater Input

The lowest power level of the highest-frequency carrier (at 128 kHz) as previously stated is –42 dBm. A carrier-to-PCM noise requirement of 57 dBm thus corresponds to a maximum allowable noise level, due to PCM interference, of:

$$\left(N_{max}\right)_{6.2\ kHz} = -42\ dBm - 57\ dB$$

$$= -99\ dBm.$$

This value is derived from the subtraction of the carrier-to-noise level (57 dB) from the lowest carrier level (–42 dBm).

or

$$\left(N_{max}\right)_{3.1\ kHz} = -102\ dBm$$

where N_{max} is the lowest allowable noise level in the 6.2-kHz or 3.1-kHz slots at 128 kHz. Dividing the bandwidth by 2 makes the requirement more severe by 3 dB and hence it becomes –102 dBm.

The maximum allowable level of PCM interference is thus 6 dB (108 dB – 102 dB = 6 dB) greater (less severe) at a low-group intermediate repeater input than at an input using the high frequency line-group. Also, the noise power in a 3.1-kHz slot centered at 128 kHz (the highest carrier frequency in the low group) of the interfering PCM signal is 6 dB lower than in such a slot centered at 264 kHz (–23 dBm compared to –17 dBm as shown in

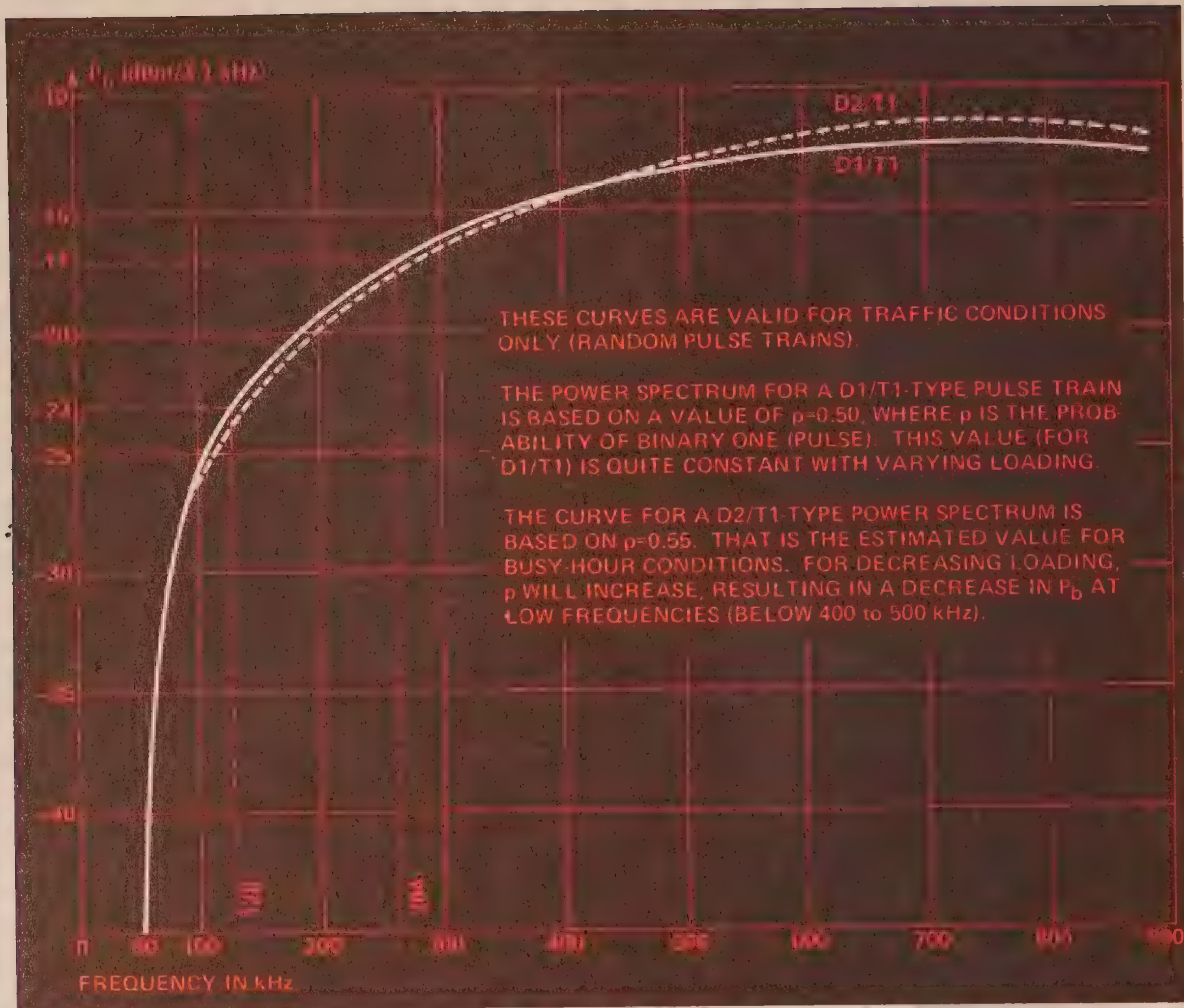


Figure 5. Power Spectrum of a T1-type PCM pulse train under traffic conditions at the output of a regenerator.

Figure 5). A low-group intermediate repeater input is thus 12 dB less sensitive to interference from a PCM system with traffic than such an input using the high line-group (due to two 6-dB advantage factors).

Based on the above calculations, it can be concluded that only the high-group intermediate repeater inputs, being the ones most sensitive to interference, are going to determine the required value of near-end or far-end crosstalk coupling loss between PCM and FDM cable pairs. Therefore, in this example the effects of all low-group repeater inputs may be neglected. The error incurred by this approximation does not exceed 0.5 dB provided that the section of the 47A system exposed to PCM interference contains more than one intermediate 47A repeater section.

PCM Noise Conclusion

For this example, in the A to B direction of transmission shown in Figure 2, two high-group intermediate repeater sections are exposed to PCM interference. Had only one such section been exposed, the near-end crosstalk coupling loss requirement would have been 91 dB at 264 kHz. Since there are two exposed high-group intermediate FDM repeater sections, the requirement is 91 + 10 log 2 dB, or 94 dB, for near-end crosstalk coupling loss at 264 kHz. This is based on one interfering PCM system and the length of its exposure to the FDM system as shown in Figure 2.

The value of 94 dB for near-end-crosstalk coupling loss is based on total PCM to FDM noise crosstalk contributions. The assumption is made here that near-end crosstalk contributions will be much more likely to cause interference than the contributions of PCM to FDM interference due to far-end crosstalk. If this assumption cannot be made, half of the crosstalk contribution can be assigned to near-end crosstalk (making that requirement 3 dB more severe, or 97 dB), and half to far-end crosstalk. The requirements for coupling losses between PCM and FDM cable pairs are then as follows:

(a) Near-end crosstalk coupling loss requirement at 264 kHz is 97 dB

(b) Far-end crosstalk coupling loss requirement at 264 kHz (as measured over one PCM repeater section length) is (97 – C dB), where C is the loss in dB at 264 kHz of the cable pair over one PCM repeater section length.

In the B to A direction of transmission shown in Figure 2, there is only one intermediate high-group repeater section exposed to PCM interference that is not adjacent to a high-noise point. For simplicity, this example has considered only the worst direction of transmission (the A to B direction).

Single-Tone Interference

The next problem to consider is that of single-tone near-end interference (between opposite directions of transmission) from the PCM system into the 47A FDM system. Figure 6 shows that the only frequency of concern in this case is 193 kHz (D1A-type signaling, and highest frequency in the FDM system of 264 kHz).

An interfering tone at 193 kHz will fall at the 1-kHz point (1 kHz on one side of the carrier) in the upper sideband of Channel 4 of the high group of the 47A system (see Channel 4, Figure 1). The maximum allowable level of such a 1-kHz tone is set to –70 dBm0, that is, 70 dB below test tone level. This is consistent with previous assumptions since this corresponds to 100 pWp at the zero-level test-tone point. The maximum allowable PCM noise level in the previous calculations was 20 dBrnc, which corresponds to 100 pW psophometrically weighted (see Figure 3). It should be remembered that the PCM to FDM interference occurs *either* as noise *or* single-tone interference, never both types simultaneously.

In this case (D1A-type signaling) only the high-group repeater inputs are of interest, since the interfering tone falls within the high-group frequency range. End sections or repeater sections adjacent to high-noise points can be disregarded as before if the length of the 47A system section being exposed to PCM interference contains at least as many normal (adjacent to low-noise points) intermediate 47A repeater sections as there are 47A repeater sections adjacent to high-noise points.

The lowest carrier level encountered at a high-group repeater input is –48 dBm, in accordance with the value used earlier in this example. This value is for the highest-frequency carrier (Channel 13) at 264 kHz. For Channel 4, the lowest level encountered is –41 dBm. Since in a 47A system without compandors the level of each sideband of a test tone is 15 dB below the carrier level (see Figure 4), the lowest such sideband level encountered is –56 dBm (–41 dBm – 15 dB). The interfering tone must be at least 70 dB below that, in other words, less than –126 dBm.

For the 47A system in this discussion, the compandor advantage allows

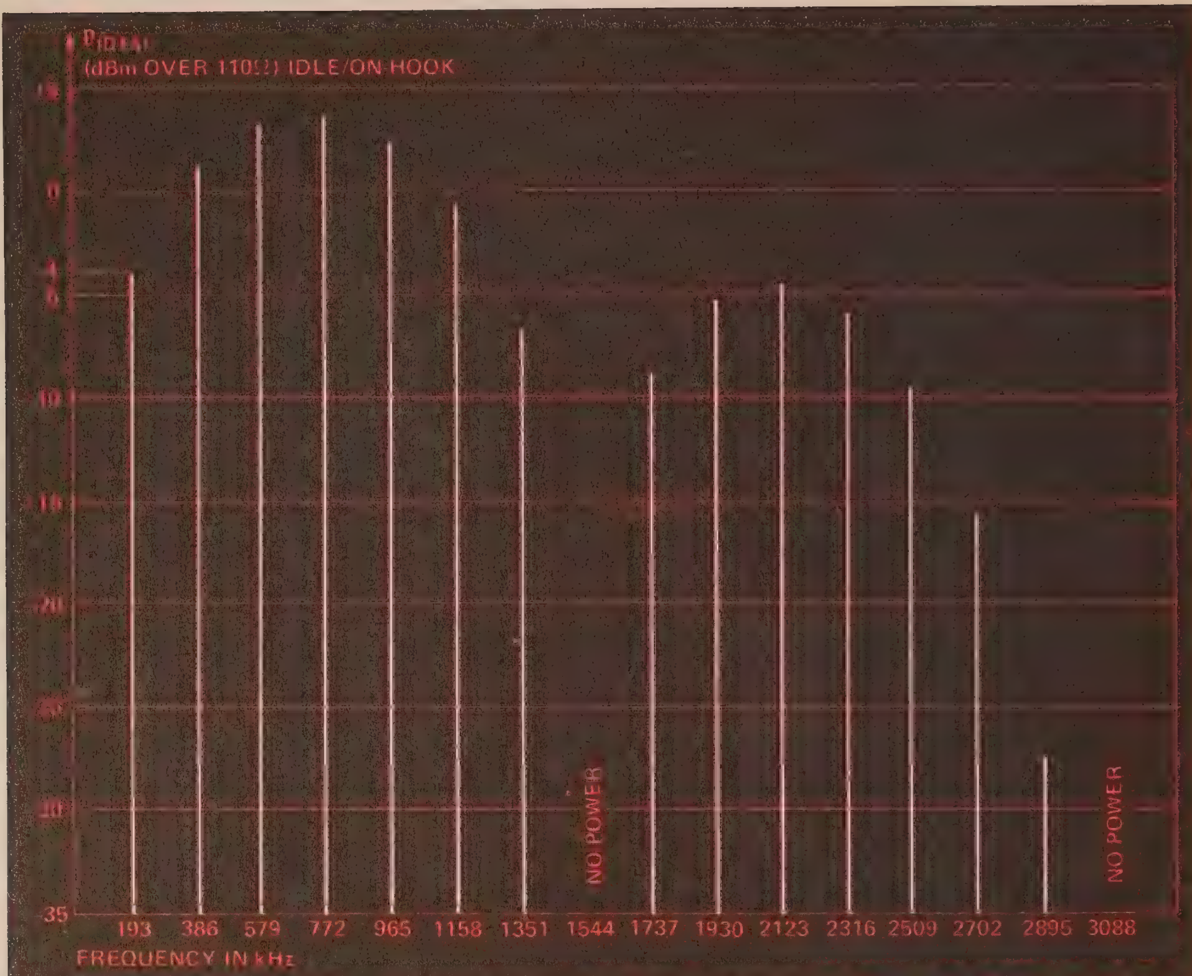


Figure 6. The line power spectrum which results in a D1-type terminal under idle/on-hook conditions when D1A (D1/D8) signaling is used.

this requirement to be relaxed by 20 dB to –106 dBm.

The level of the 193-kHz tone of an idle PCM system with D1A-type signaling is –4 dBm as shown in Figure 6. The difference between –4 dBm and –106 dBm is 102 dB. The requirement for near-end crosstalk coupling loss at 193 kHz between PCM and FDM cable pairs thus becomes 102 dB per exposed high-group 47A intermediate repeater section. Since there are two such sections exposed in the A to B direction of transmission as shown in Figure 2, this requirement becomes 105 dB (102 + 10 log 2 dB = 105 dB) for this direction. (The requirement for the other direction is 3 dB less, but for simplicity only the worst case will be considered here.) This 105-dB value is based *only* on single-tone interference and on the assumption that the following additional circumstances are true:

(a) the effect of far-end PCM to FDM crosstalk can be neglected

(b) only one PCM system is involved.

If far-end PCM to FDM crosstalk can for some reason not be neglected, 3 dB should be added to the requirement above. This new value (108 dB at 193 kHz) becomes the new near-end crosstalk requirement; (108-D) dB is the far-end crosstalk requirement as measured over one PCM repeater section length. (D is the loss in dB of the cable pair at 193 kHz over one repeater section length of the PCM system.) This assumption effectively assigns

half of the single-tone interference to near-end crosstalk and half to far-end crosstalk contributions.

Overall Conclusion For This Example

In this example, the resulting requirements on near-end and far-end crosstalk were:

(a) *For near-end crosstalk over one repeater section length of the PCM system,*
97 dB at 264 kHz
108 dB at 193 kHz

(b) *For far-end crosstalk over one repeater section length of the PCM system,*
(97-C) dB at 264 kHz
(108-D) dB at 193 kHz

where, C and D are the losses in dB of the cable pair at 264 and 193 kHz, respectively, over one PCM system repeater section length.

If the requirements at 264 kHz can be met (this is necessary to combat noise across the band coming from the PCM system when it is loaded with traffic) but the requirements at 193 kHz cannot be simultaneously met, consideration should be given to removing the 47A channel affected (Channel 4) from service.

Conclusions Regarding "T-to-N" PCM-FDM Interference

The results and conclusions of the example just discussed are valid for any D1/T1-type PCM system equipped for D1A-type signaling, disturbing an N1 or N2-type FDM system equipped with compandors.

If the D1/T1 system is equipped for D1B signaling there will be an additional component of single-tone interference falling at 96.5 kHz, corresponding to the 500-Hz point (on one side of the carrier) of one of the sidebands of Channel 6 in the lower line group. This will produce an additional requirement (at that frequency) for near-end and far-end crosstalk coupling loss, respectively.

If the PCM system is a D2/T1-type system, the single-tone interference problem is reduced significantly.

The calculations regarding noise interference from a traffic-loaded PCM system are the same for a D1B as for a D1A-type PCM system. For a D2/T1-type system, a separate curve is used (see Figure 5).

If a D1/T1- or D2/T1-type PCM system is disturbing an N3-type system such as the GTE Lenkurt 46B, there is a 3-dB disadvantage since the single-sideband 46B system does not have the advantage of coherent detection of a double-sideband signal (as the 47A does) since,

Equation 2

$$\left(\frac{P}{N}\right)_{6.2\text{ kHz}} = \left(\frac{S}{N}\right)_{3.1\text{ kHz}} + 12\text{ dB}$$

where P/N is the pilot-to-PCM interference noise ratio, and S/N is the test tone-to-PCM interference noise ratio. In equation 2, the 3-dB disadvantage shows up in the +12-dB value when compared to the +9-dB value of Equation 1.

This three part series on PCM-FDM compatibility has endeavored to provide the user of communications equipment with as much information as has been gathered to date by GTE Lenkurt on the problem of combining PCM and FDM within the same cable sheath. Adherence to the ground-rules laid out in this series should enable the user to systematically phase out an old FDM system while using PCM, or permanently combine PCM and FDM within the same cable sheath.

GTE LENKURT

DEMODULATOR

FEBRUARY 1973

The PCM repeatered line is a system of physical wire lines with repeaters situated at specific locations along the way. It is responsible for accurately reproducing the digital signals that are transmitted between two PCM terminals.

Legend has it that in the early days of telephony, the original repeaters had human form. It seems that before electronic repeaters were developed to amplify voice signals along a telephone line, a human being (called a repeater) would periodically relay information along the line, from the caller to another repeater, or to the final listener. Later, vacuum tube amplifiers replaced the human repeater and adopted his name. Then, solid-state repeaters, were developed to amplify voice frequencies along a line. With the emergence of PCM (pulse code modulation), a totally different type of repeater was developed to accommodate the totally different PCM transmission medium.

PCM Channel Banks

PCM terminals (channel banks) originate and receive the digital information which is conveyed along the PCM repeatered line. There are two types of terminals which are in general use – the D1- and the D2-type. Both D1- and D2-type channel banks can function over the T1-type repeatered line (which is the subject of this discussion), and both sample each voice channel 8000 times per second. Each can handle 24 voice channels over two pairs of cable (one pair for each direction of transmission). One of the major differences between these two terminals is that the D1-type has 127 discrete voltage levels available for quantization, while the D2-type has 255 available levels. The greater number of available levels of the D2-type terminal results in a more accurate representation of the original voice signal at the distant terminal. (See the August, 1971 and January, 1972 issues of the Demodulator for a discussion on D1- and D2-type PCM channel banks.) The digital pulse train generated by the PCM terminal and processed by the repeater is in bipolar form. This means that each pulse (or binary "one") is directly opposite in polarity to the previous binary even if there is a string of zeros in between (see Figure 1).

PCM Repeaters

The purpose of a PCM repeater is to construct an accurate reproduction of the original digital pulse train as it originated at the PCM terminal. The three major functions involved in this process are reshaping, regeneration, and retiming.

The original design of PCM repeaters had to take into consideration existing cable parameters, since economics dictated that existing cable plant be used. In voice frequency transmission systems, loading coils are installed along the line according to a predetermined loading plan, to increase line inductance and thereby improve transmission characteristics. For example, an H88 loading plan requires that 88-millihenry coils be placed at 6,000-foot intervals (except at end sections where spacing is 3,000 feet). Also, if a loading coil cannot be installed at the predetermined 6,000-foot interval, line-building-out networks must be installed to make up the distance required by the loading plan. Both loading coils and line-

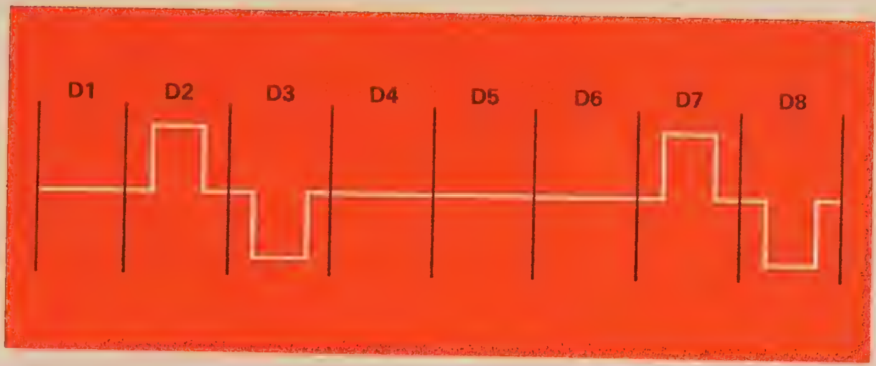


Figure 1. In bipolar format, each pulse is of opposite polarity from the preceding one, even if there is a string of zeros in between.

building-out networks must be removed from existing voice-frequency cable plant if it is to be used for any carrier systems above 4 kHz.

Taking into consideration the 6,000-foot intervals of loading coils, PCM repeaters were designed to faithfully reproduce digital pulses at distances of 6,000 feet or less. This is true for 22-gauge, paper insulated, 0.083 μf cable. The repeater itself was designed to reproduce pulses which have a maximum attenuation of approximately 35 dB at 772 kHz. With these transmission design characteristics, repeaters could be placed at the same location previously occupied by the loading coils. This means that a repeater is necessary for every 6,000 feet (approximately 1 mile) of 22 gauge cable. However, depending on the gauge and capacity of a cable, repeaters may be spaced at up to 2-mile intervals. The advantage of a PCM system, as well as other types of cable-carrier systems, is that for each two cable pairs which previously accommodated two telephone channels, 24 two-way voice channels can now be accommodated.

Reshaping (Equalization)

Reshaping of incoming pulses is the first of three major functions that take place in the repeater. A pulse at a point 6,000 feet from a terminal or repeater will be distorted and may appear as shown in Figure 2. Since the pulse is spread out in time, it is the task of the repeater equalizer to restore the pulse to its original width. It should be noted that it is not necessary or even desirable to restore the pulse to a perfectly rectangular configuration. Hence, the restoration results in a limiting of the frequency band and the production of a rounded pulse at the output of the equalizer. A reshaped pulse might appear as shown in Figure 3.

Regeneration

After the reshaping function, the pulse train is amplified to a predetermined value before proceeding to the regeneration function. The pulse train then enters a voltage-comparator circuit which has a threshold such that if any incoming pulse does not exceed 50% of the nominal pulse height, there

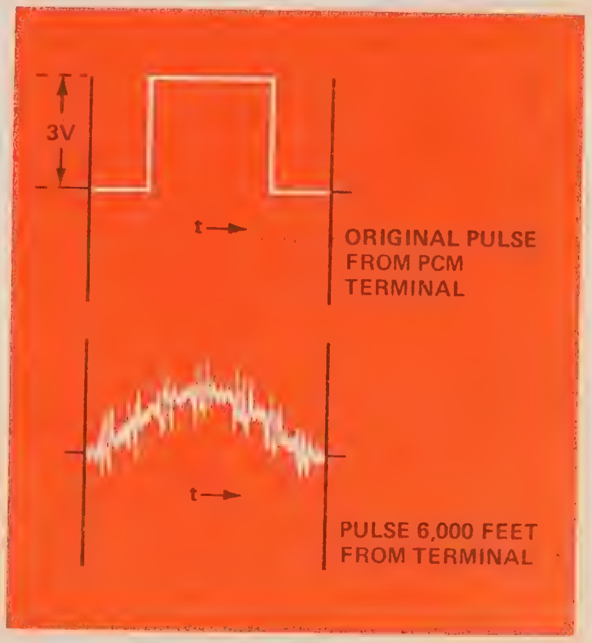


Figure 2. The original pulse is distorted in shape, amplitude, and time, as it travels along the line.

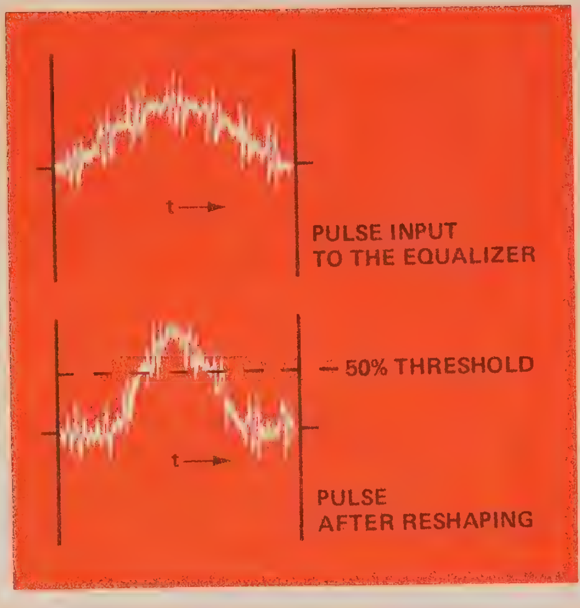


Figure 3. The reshaping function restores the correct pulse width to the incoming pulse.

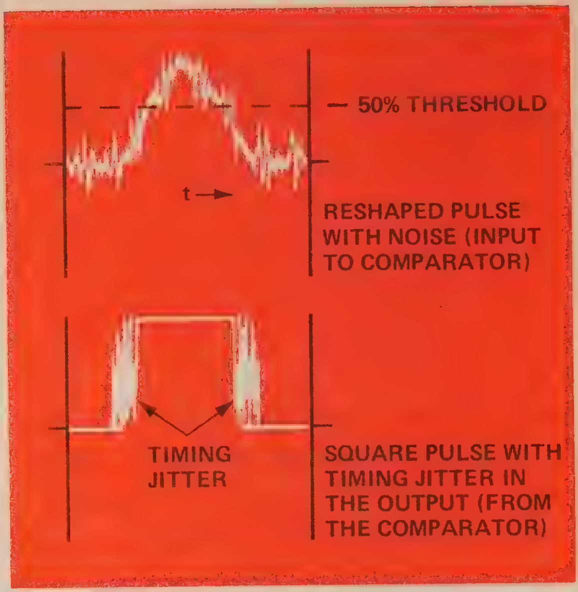


Figure 4. The voltage comparator circuit detects the presence or absence of pulses.

will be no change in the output state from the comparator, for that pulse. Essentially, this circuit is used to detect whether there is a pulse present. If there is no output change from the comparator, a zero (no pulse) condition is assumed. A pulse train may have frequency distortion, amplitude distortion, or timing distortion. All these forms of distortion must be minimized by the repeater functions. In the comparator, the level of the incoming pulse plus the noise riding on the pulse must add up to more than the comparator threshold if there is to be a one (pulse) output. An output pulse from the comparator is a square wave with timing jitter, and might appear as shown in Figure 4.

Retiming

The digital pulse train of a PCM system carries its own timing information. The beginning of the timing process takes place after the pulse has been equalized in the reshaping circuit, amplified, and brought to the correct level in the regeneration circuit. The signal path is then split in two directions – one direction goes into the regeneration circuit, and the other into a full-wave rectifier (see Figure 5). The rectifier transforms the bipolar pulse train into a unipolar pulse train so that each pulse has the same polarity. This unipolar pulse contains the timing wave or clock which must be extracted to provide the repeater timing. Extraction is accomplished by passing the unipolar pulse train through a bandpass filter (LC tank circuit), to produce a 1.544-MHz sine wave. The output from the bandpass filter is fed to an amplitude limiter whose output will be a constant-amplitude square wave. This square wave is differentiated, and produces the appropriate timing spikes, which are fed to the timing circuit to provide the correct timing information for the repeater.

The clock (LC tank circuit) must be triggered at least two out of every 16 pulse periods (approximately 10 μs) to maintain oscillation. To accommodate this, the channel bank suppresses any all-zero codes, thus insuring a minimum of one pulse every word time (approximately 5 μs).

The retiming function must correct the presence of timing jitter in the

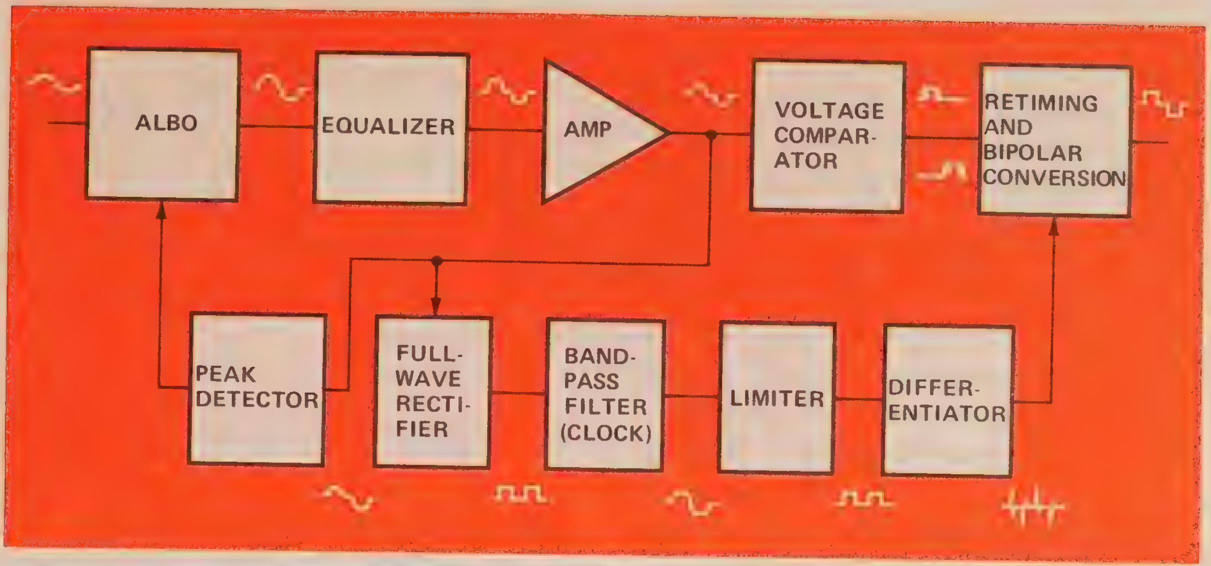


Figure 5. The digital pulse train supplies its own timing information.

pulses coming from the comparator. To do this, the retiming circuit is designed to produce a sample pulse which must be lined up with the nominal center of the pulse coming from the comparator, to minimize the effects of timing jitter present at the pulse edges. A logic function "looks" at the nominal center, and detects if a pulse is present. If a pulse is present, a new pulse is started in the pulse output circuit. The clock also generates a turn-off command which terminates the new pulse (see Figure 6). The only condition that must be met to generate a pulse is to have a "one" output from the comparator when the sample pulse occurs.

One of the problems in the reshaping function is to assure that the pulse, which has widened in time, does not overlap into the next time frame and interfere with the detection of a no-pulse condition. So, one of the functions of pulse reshaping is to make sure that the reshaped pulse dies out by the time the next point in time appears when a measurement (sample) must be taken.

In the regeneration circuit it is important that the threshold be accurately established. And, in the retiming function, it is important that the timing take place in exactly the correct region, since a pulse carrying enough noise to tax the limits of the repeater will have considerable jitter at the output of the regenerator (comparator), as in a worse-case condition. There will be a very narrow region in time where a certain output dictates a pulse condition, but once this condition is detected, a new, noise-free pulse is produced by the repeater.

Repeater Limitations

In theory, an indefinite number of repeaters can be placed in a PCM

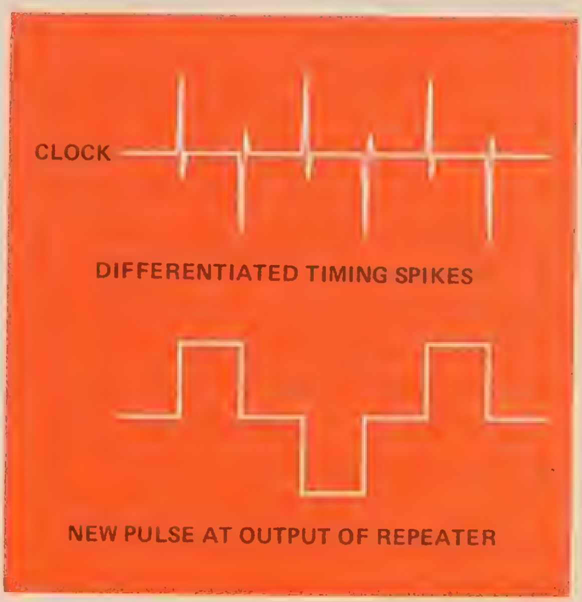
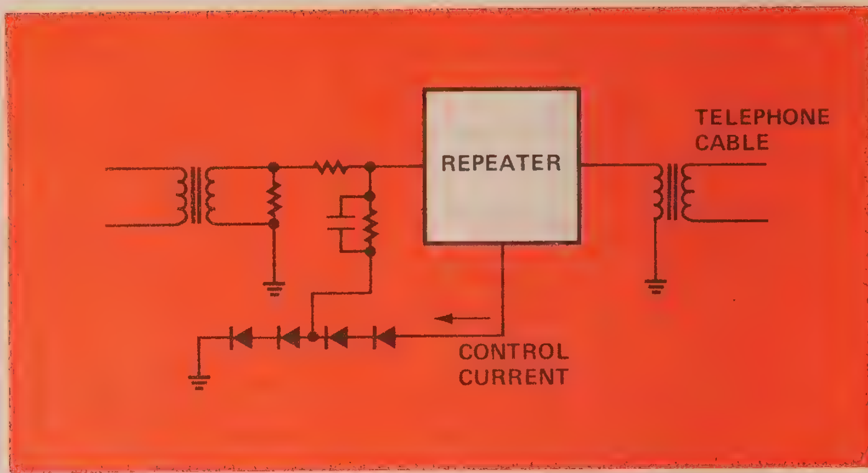


Figure 6. The timing spikes begin and terminate the new pulse.

Figure 7. The automatic line-building-out network maintains a constant electrical length on a section of line.



system; in actual practice this is not the case. Economic limitations prohibit electronic perfection in each of the repeater functions. A slight amount of deterioration is allowed in each repeater, mostly in the form of timing jitter. This means that the pulse train leaving a repeater is not perfectly retimed. The accumulation of timing jitter as the pulse train progresses from one repeater to another slowly reduces the margin-to-noise ratio. Also, the pulse train may be disturbed by crosstalk from other systems or by impulse noise in the cable which comes from relay systems that might be switching in the same cable. These forms of interference influence the clock to a certain extent and result in timing jitter. Because the timing is recovered from the incoming pulse train itself, signals carrying jitter from a previous repeater arithmetically transfer their jitter to the next repeater. It is possible to eliminate most of this accumulated jitter by periodically inserting a special dejittering device, but this would require a very expensive repeater with a highly accurate clock. This might be done in a case where it became necessary to extend the present length of a system.

Automatic Line-Building-Out Network

The major development in PCM repeaters since their original design is the automatic line-building-out network (ALBO). The first designs were such that if the spacing between repeaters wasn't close to 6,000 feet, small artificial lines had to be installed at the input of the repeater. Variations in repeater spacing are now controlled by an automatic circuit. The peak of an incoming pulse is detected by a peak detector which then feeds a control signal into a control element so that the same peak value is maintained at all times, regardless of cable length, as long as the length is less than 6,000 feet. As the cable length varies, the ALBO readjusts to maintain what appears to be a constant length going into the equalizer of the repeater.

The operation of the ALBO is relatively simple. The device, appears schematically as shown in Figure 7. Part of the ALBO is a variable resistance which is realized by varying the current in silicon diodes. This control comes from the repeater, and is proportional to the peak of the input signal. If a cable pair is very short, the signal-input to the repeater will be strong. The peak detector senses the strong signal and puts out a large current which makes the diode resistance very low, thus producing a large shunting affect on the signal. If the diode resistance is low, there is an increase of attenuation to the network. In this way the level of the signal at the input to the repeater

always has the same value. Two cable pairs are required for two-way PCM telephone communication. Each repeater, then, has two separate devices within, each capable of handling the digital bit stream in one direction. And, each device is independent of the other except for a common power supply.

Cable

Although there have been some improvements in voice frequency cable that yield better transmission characteristics, no significant changes have been made. Some recent efforts have also been made to place two sets of groups of pairs within the same cable sheath, separated by a shield, to eliminate near-end crosstalk problems. There is work going on now to develop cables with internal shielding for T2-type repeatered lines which must operate at more than four times the speed of the T1-type. Another alternative is to use separate cables for each direction of transmission. This would yield better transmission characteristics but would be the more expensive alternative. The T2-type repeatered line will have a capacity of 96 voice channels on two cable pair, and, if special low-capacity cable is used, spacing between repeaters will be around 12,000 feet – twice as long as for T1 line.

Cable Testing

As mentioned earlier, the bulk of PCM systems being installed today are taking advantage of existing voice-frequency cable. There is, however, a significant difference between the transmission of a square, digital pulse and the transmission of a sinusoidal voice frequency. A cable pair, for example, may work perfectly for voice frequencies, but be completely inoperative when PCM information is placed on it. This has to do, in part, with the frequency band over which the equipment is operating. The square wave has an infinite number of frequencies in the form of odd harmonics in its composition, while with a sinusoidal voice frequency, only the fundamental frequency must be transmitted down the line. It is therefore important for a user of telecommunications equipment to be able to determine which of his existing cables are suitable for PCM transmission.

There are several ways of testing voice-frequency cable for PCM; one of the simplest involves the use of a test set recently developed by GTE Lenkurt. This test set, called the GTE Lenkurt 91100 PCM Cable Pair Test Set, can make three major evaluations, each of which involves the transmission of a *true* PCM signal from the test set. The test set can measure the approximate attenuation of a PCM signal in dB, on any cable pair between repeaters. It can also measure the margin of degradation in a cable pair. This measurement not only tells the user if he has a cable pair suitable for PCM, but also tells him how much more interference a cable pair can tolerate before it can no longer be used for PCM. The third use for the test set is to check a PCM system from end-to-end. This measurement is used when it is uncertain if a transmission problem lies in the terminals or in the repeatered line.

The move toward PCM systems has been gaining momentum for some years. It is one of the major transmission mediums of the future. Already plans are being conceived which make almost exclusive use of PCM transmission. PCM over microwave is also now out of the experimental stage so that in the future, PCM signals will not only travel through cable, but also through space.

GTE LENKURT
DEMODULATOR
FEBRUARY 1974

EXTENDED LENGTH PCM SYSTEMS

Since PCM carrier systems were introduced several years ago, the recommended maximum length of the 1.5-Mbs, T1-type repeatered line has been held conservatively at approximately 50 miles. Recent calculations and field measurements indicate that this length can be extended to 200 to 300 miles.

The first PCM systems were designed for use as metropolitan, toll-connecting type of carriers. Today, with a need for longer PCM systems, the previous 50-mile PCM line limit is getting a second look by manufacturers of telecommunications equipment.

Recently, telephone operating companies have found it advantageous to design much longer exchange or toll-connecting trunks in rural areas. This has come about as a result of the increasing popularity of PCM systems. Thus PCM is being considered for expansion on routes being served by open wire, older cable carrier systems, or even radio. While this is most prevalent in the more sparsely settled areas, the concept is certain to spread. This has made it necessary to establish the limits by which the proper functioning of a PCM system is bound. In this light, new calculations have been made and experiments conducted by GTE Lenkurt, in an effort to determine the maximum number of PCM repeaters that may be placed in tandem, while still retaining proper operating conditions. Some factors which were considered as possible limiting elements were: (1) powering characteristics, (2) fault location, and (3) phase jitter. The powering characteristics are mainly the problems connected with bringing dc power to the repeaters. This did not prove to be a limiting factor, since power stations are usually available along the line. Location of faults along an extended line was also found not to be a limiting factor. In the final analysis, it was found that phase jitter was the limiting factor to the establishment of extended PCM systems.

Phase Jitter

Phase jitter is an abrupt variation in the phase of the PCM signal, and may be caused by such things as impulse noise, noise from other systems, and pulse pattern changes in the digital bit stream (see Figure 1A). At the input of a repeater, the incoming bit stream is fed into an LC tank circuit. This causes an energy transfer such that the clock can be kept going at a fixed rate, even if there is a series of zeros in between. It is necessary that the incoming pulses and the clock pulse occur at the same time. Phase jitter can offset a PCM pulse far enough that it may not be detected by the clock pulse, thus causing an error in the system.

The timing or clock for the repeatered line is recovered from the incoming PCM signal. This clock is used to generate sampling spikes which should properly sample at the center of the incoming pulse, as shown in Figure 1B. It might be surmised that if the clock stays in the same position, then there is less margin of detecting a pulse, because of the phase jitter on that pulse. This is true in the case of high-frequency jitter (jitter caused by impulse noise, or noise from other

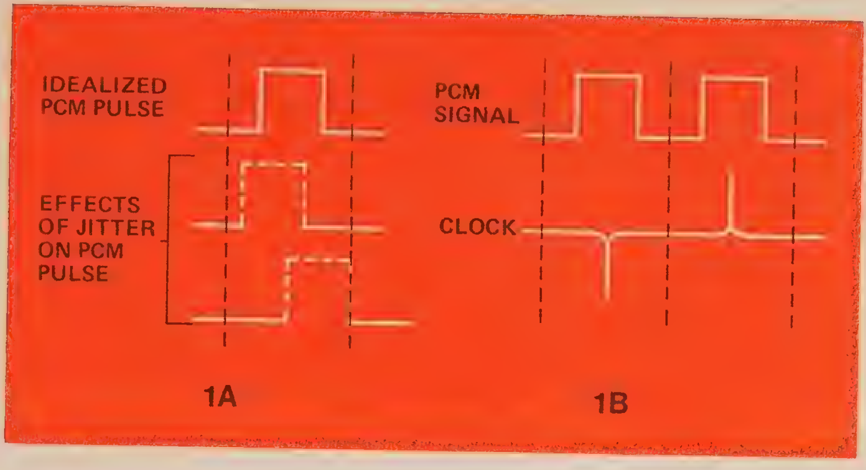


Figure 1A. Phase jitter causes a variation in the phase of the PCM signal.

Figure 1B. Clock timing pulses should sample at the center of the incoming pulses.

systems). But generally, what passes on down the repeatered line is low-frequency jitter, which is the most common type of phase jitter and which is mainly caused by changes in the pulse patterns of a system. Since pulse patterns are constantly changing in a PCM system, there will be a constant source of low frequency jitter. However, this type of jitter is relatively tolerable in an extended length PCM system, since the clock moves back and forth with the change in pulse pattern. This is possible because the clock is a regenerated one, rather than a crystal-controlled internal clock. The result of this is that the clock pulse will sample the incoming pulses at their center, as is desirable. (See the February, 1973 issue of the Demodulator for further information on PCM repeatered line operation.)

Multiplexers and Jitter

Basically, what jitter does is change the instantaneous frequency on the line to some frequency above or below the nominal frequency. According to recent calculations and experiments, it has been found that with stable repeaters, phase jitter on the repeaters does not place a limit on the length of the line between terminals. However, phase jitter does place a limit on the length of a PCM line when two or more lower speed PCM lines are multiplexed to one higher speed PCM line.

It is actually the multiplexing that limits the length of the PCM line. At the input to a PCM multiplexer, only a limited range of frequency change above or below the nominal line frequency of 1.544 Mbs can be tolerated. If jitter causes a frequency outside that range, then the multiplexer will cause errors.

When a pattern shift occurs, it will shift the pulses back and forth approximately two nanoseconds per repeater. And, because there are a series of regenerative repeaters in a PCM line, each one of them shifts the same amount. Recent tests have shown that two PCM channel banks, with stable repeaters, such as the GTE Lenkurt 9101C, can operate practically as well with 500 or more miles of repeatered line separating them as they can when operating back-to-back. However, the limit of the length of the repeatered line has been shown to be about 200 repeaters, if a multiplexer is to be used. Since many PCM systems will eventually be used in conjunction with multiplexers, a maximum of 200 repeaters is recommended between terminals, so that an eventual transition to a multiplexed system may be simplified. It should be noted that jitter considerations limit the number of repeaters on a T1 line, and not the length of the line in miles. Thus, with 5360-foot spacing between repeaters, 200 repeaters could extend for 200

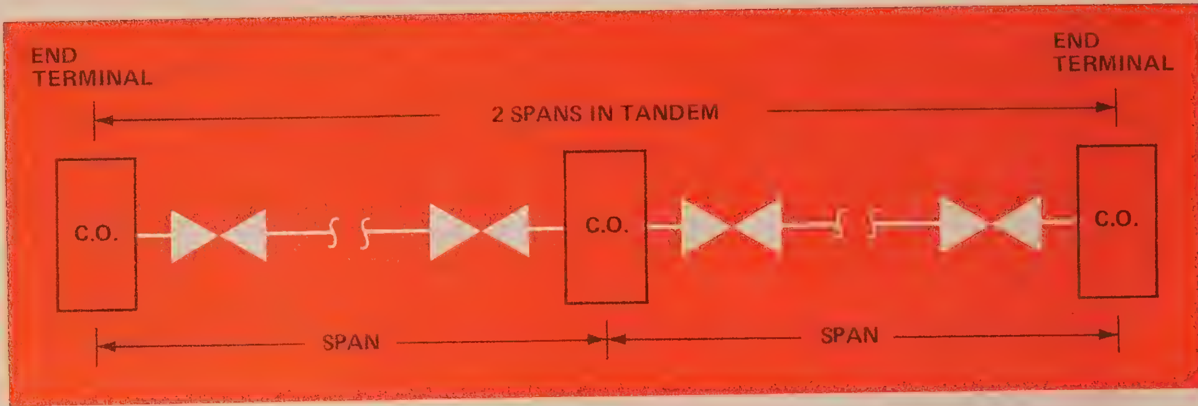


Figure 2. As it becomes necessary to increase the length of repeated lines, several spans can be placed in tandem.

miles, but with 7900-foot spacing, 200 repeaters could extend for 300 miles.

Limitations in Repeater Spacing

As longer repeated lines are needed, several "spans" will necessarily be place in tandem. (A span includes all the repeaters between two central offices, which provide powering and maintenance facilities, see Figure 2.)

The maximum repeater spacing within a span is determined by the allowable error rate. This error rate, in turn, is determined by the number of tandem spans, type of cable, placement of transmission pairs within the cable (crosstalk coupling loss), and proximity of repeater pairs to a central office. The total error rate between two end terminals separated by several tandem spans is essentially the sum of the individual span error rates.

The maximum error rate between two end terminals providing good voice communications is generally accepted to be 1×10^{-6}. The error rate is the total amount of information in error that is caused by the transmission media, divided by the total amount of information received. An error rate of 1×10^{-6} would mean that there is one error in one million units of information. A common error rate objective per span has been 3×10^{-7}. Thus, if all span lines were to operate at their maximum error rate, only three spans in tandem could be allowed in a system.

To determine if repeater spacing should be more conservative with many spans in tandem, repeater spacings were calculated for several cable types with 1, 3, and 10 spans in tandem.

A criterion was used which calls for a reduction in error rate per span as tandem spans are added, so that all spans can operate at this error rate without the overall error rate exceeding 1×10^{-6}. Thus, for one span, the allowed span error rate is 1×10^{-6}; for 3 spans in tandem, the allowed error rate is 3×10^{-7} per span; and for 10 spans in tandem, the allowed error rate is 1×10^{-7} per span.

The allowed error rate per span is further apportioned over each span by allowing each of its two end sections (between an office and first repeater) to have 1/3 of the total span error rate and by allowing the section between the first and last repeater in a span to have the remaining 1/3 of the total span error rate.

Once these maximum span section error rates have been calculated for a fixed length system, the maximum repeater spacing allowable to achieve these error rate objectives can be calculated based on the limiting source of line interference. In one-cable operation, near-end crosstalk is the most

serious type of line interference encountered on the midsections of a span. In two-cable operation, far-end crosstalk is the most serious type of line interference encountered on mid-sections. In both one and two-cable operation, end-section line interference is due to office impulse noise.

Although each cable is different, a typical result showing maximum mid-section repeater spacing for one of the cable types considered, is shown in Figure 3. This shows that mid-section repeater spacing with up to 10 spans in tandem does not need to be significantly more conservative than when calculated using the present standard method (i.e. span error rate of 3 x 10^{-7}, or 3 spans in tandem). For several types of cable considered, the mid-section spacing for 10 spans in tandem was only between 100 and 150 feet shorter than for 3 spans in tandem. However, more conservative end-section repeater spacing is required with multiple tandem spans. Figure 4 shows the end section of a repeatered line.

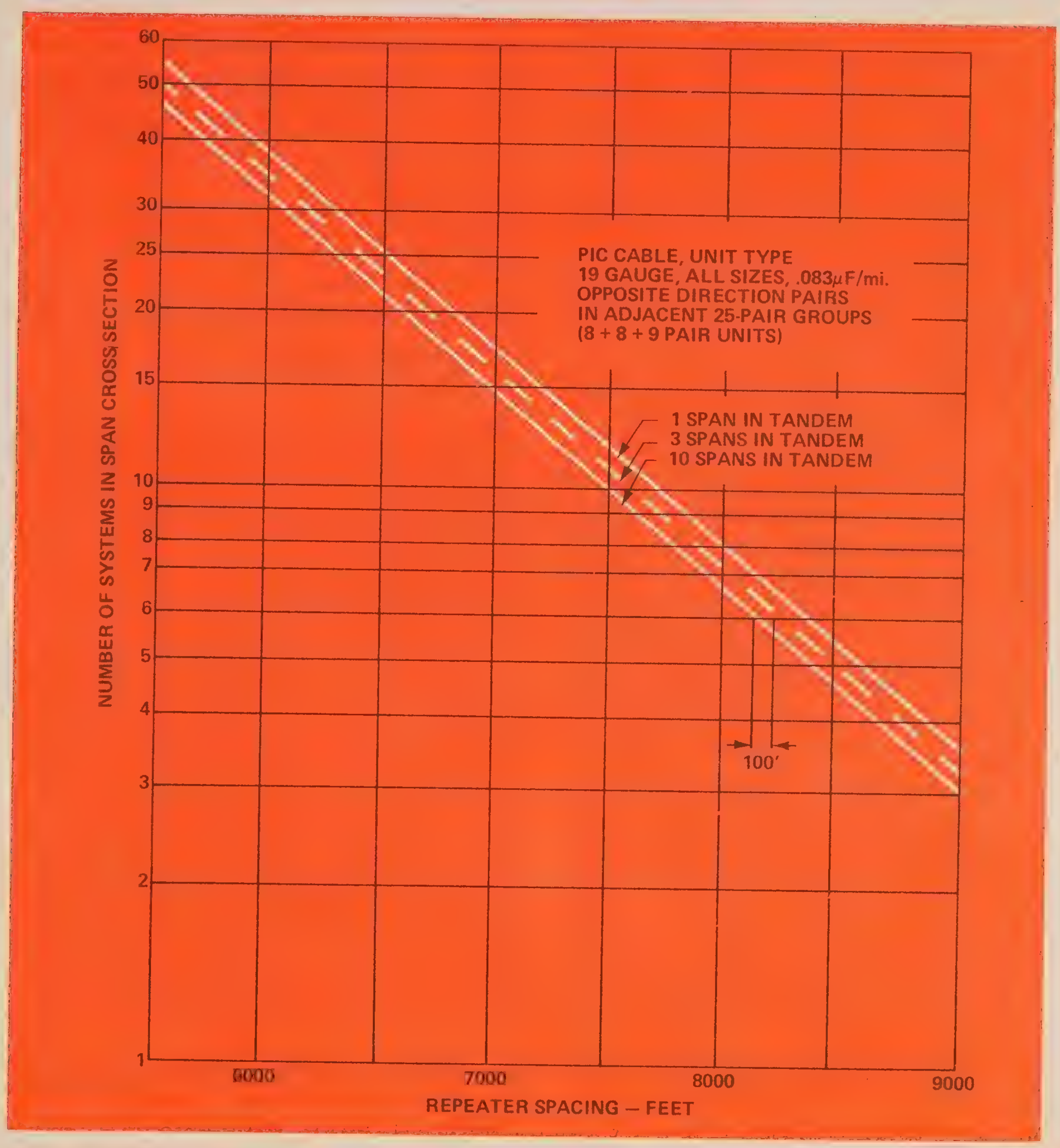


Figure 3. Mid-section repeater spacing with up to 10 spans in tandem does not differ significantly from present standards.

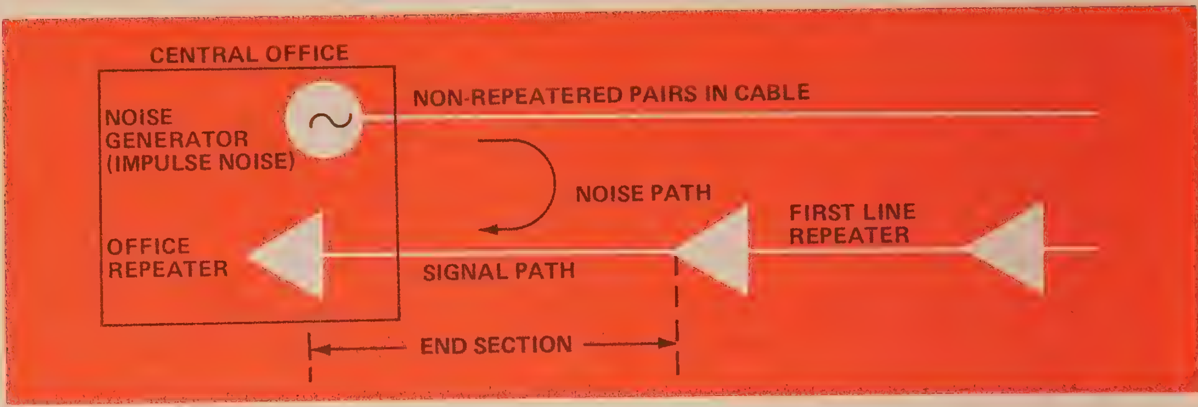


Figure 4. More conservative end-section repeater spacing is required with an increase in tandem spans.

The signal-to-noise ratio and therefore the error rate of an office repeater, is governed by the difference between noise path loss and signal path loss in the end-section of the repeatered line (see Figure 5). With the noise path loss arbitrarily taken to be 75dB for nearly any cable case, the allowable end section loss can be calculated based on the desired error rate. Using the graph of Figure 5 for error rate vs. difference in loss of signal and noise path, the allowable end section loss can be calculated for various error rates. As the number of tandem spans increases, the lower allowed end section error rates cause the end section repeater spacings to be shortened (see in Figure 6). For example, to attain an error rate of 1×10^{-7}, the difference would be approximately 52dB. Since the voice path loss is assumed to be 75dB, then 75dB – 52dB equals an end-section loss of 23dB. Using the same calculation for an end section error rate of 5×10^{-8}, the end section loss would be, 75dB – 55dB = 20dB.

Administrative considerations

With a repeatered line length limitation of about 200 repeaters, the number of spans could become very large (as well as the corresponding number of intermediate offices).

There are several administrative and service quality considerations which must be kept in mind when planning for many spans. These are mainly management problems and are more closely related to the number of intermediate offices in the complete line rather than to the number of repeaters.

Some of these are:

(1) the chance of excessive "down time" during a span failure while the faulty span is located and the system is patched to a spare line

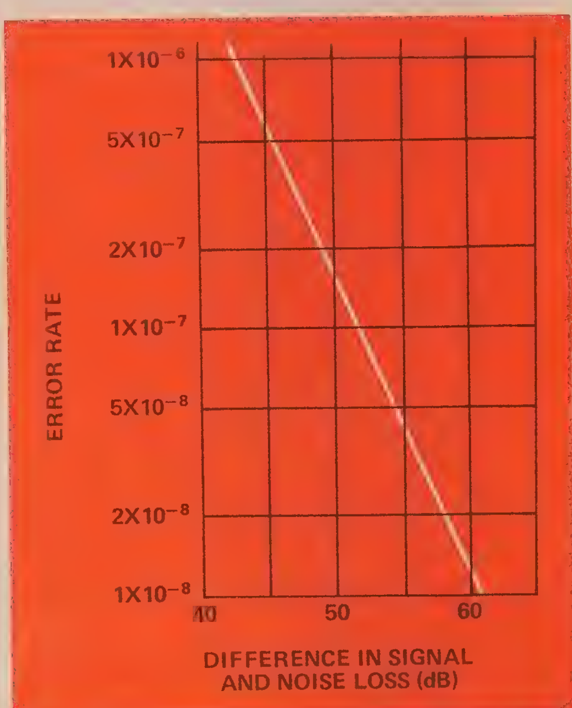


Figure 5. The error rate of an office repeater is governed by the difference between noise path loss and signal path loss in the end-section of a repeatered line.

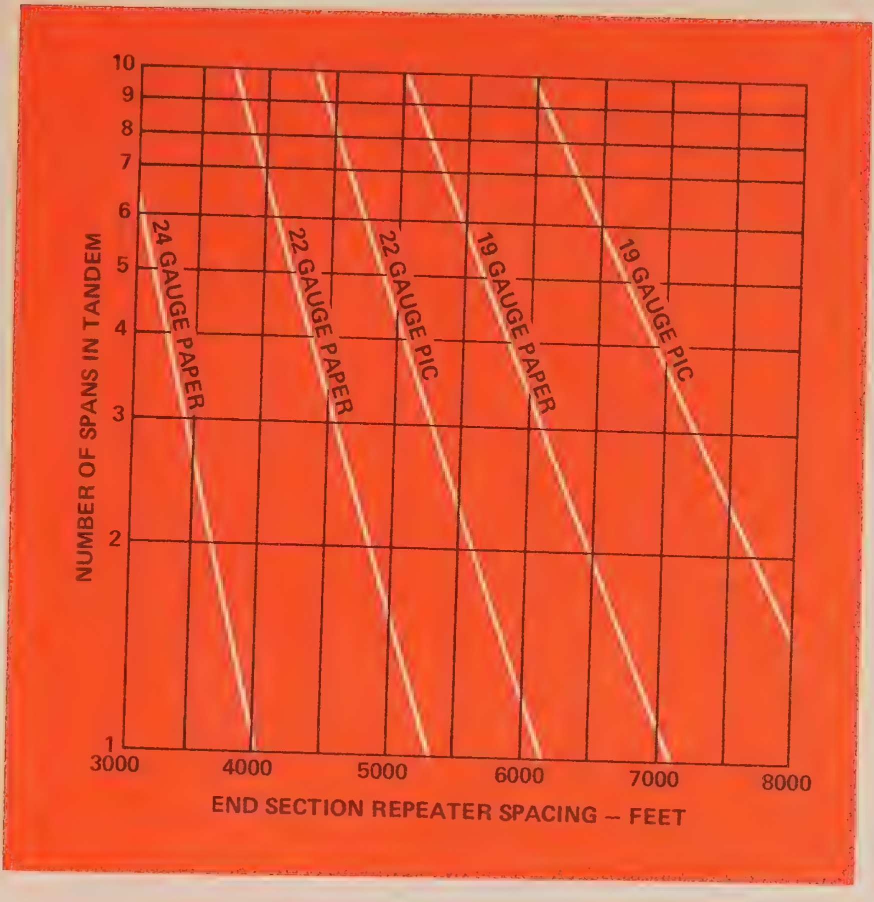


Figure 6. Maximum allowable end-section repeaters spacing for various numbers of spans and cable sizes.

(2) spans have to be divided because of administrative differences in central offices or because of area boundaries

(3) accurate record keeping becomes increasingly difficult as more offices are added, where each office requires a complete set of records for the systems entering and leaving it

(4) the possibility of human error causing system failures increases as intermediate offices (and particularly patching jack facilities) are added

With these and other considerations, it is well to keep the number of intermediate offices to a minimum in a long system, so that the administrative considerations will not impose an undue burden on system lengths.

BIBLIOGRAPHY

1. Byrne, C. J., B. J. Karafin, and D. B. Robinson Jr. "Systematic Jitter in a Chain of Digital Repeaters," *Bell System Technical Journal*, Vol. 42, (November 1963), 2679-2714.

2. Cravis, H. and T. V. Carter. "Engineering of T1 Carrier System Repeatered Lines," *Bell System Technical Journal*, Vol. 42, (March 1963), 431-486.

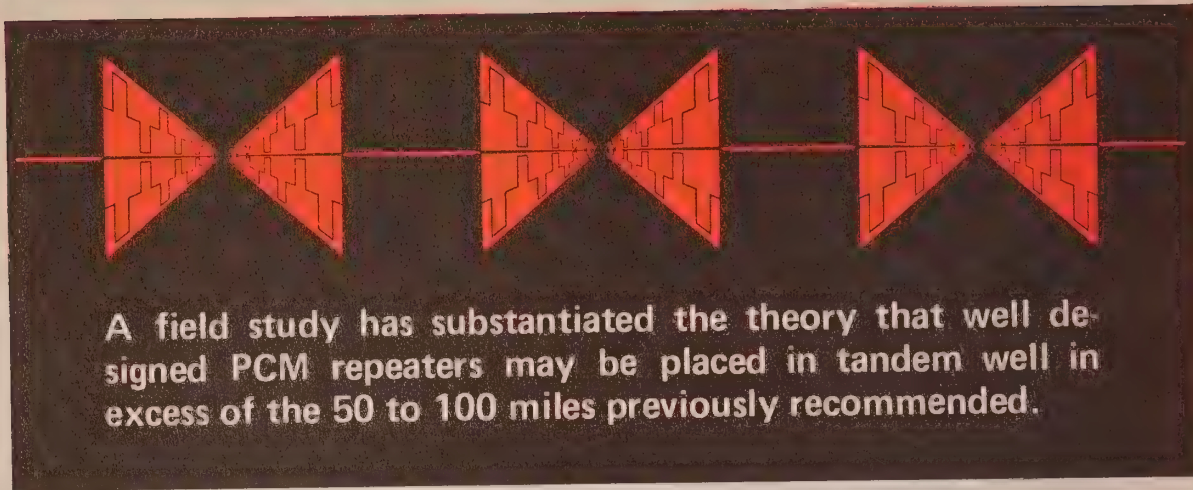
3. Crawforth, L. D. and J. D. Olson. "Maximum T1 Repeatered Line Length Considerations," *International Conference on Communications*, Vol. II, (June 1973), 32-28 – 32-33.

GTE LENKURT

DEMODULATOR

MARCH 1974

PCM Field Study



Theory indicates that the length of a PCM (pulse code modulation) repeatered line is not limited. But, when several low-speed PCM lines are to be multiplexed into one high-speed line, low-frequency phase jitter, which arises when pulse-pattern changes take place, does limit the line to about 200 repeaters. Since phase jitter has the effect of momentarily speeding up or slowing down the pulse stream, errors will be created in the output bit stream of a PCM multiplexer if the elastic store of the system either overflows (too fast for too long) or empties out (too slow for too long). (See the *GTE Lenkurt Demodulator* for February, 1974.) Accepting these theories, a field study was undertaken to substantiate them.

The particular span that was available for this study afforded an opportunity to build up a tandem system of 650 regenerators (repeaters), extending for 1000 miles. Until this time, no published data was known to exist for systems in excess of about 100 miles, and available recommendations on system design traditionally limited the lengths to between 50 and 100 miles. The opportunity to measure performance in an actual system containing more than 200 regenerators was unique, both as a validation of the theoretical results and as a contribution to new design philosophies for T1-type carrier systems.

Test Line

The span to be used for this study was a dedicated T1 carrier facility between two class 5 offices, twenty miles apart. The cable used was new, 19-gauge, gel-filled, polyethylene-insulated, D-screen cable. It was totally buried except for the end sections, which were in conduit. Two sets of repeaters—one for each direction of transmission—were located in manholes, in pressure-type housings. The repeater spacing, which averaged 8900 feet, allowed for temperature build-up and for the anticipated 100% fill of the cable. Figure 1 shows the repeatered line layout used for the field study.

PCM repeaters, such as the GTE Lenkurt 9101C, are actually two regenerators in one enclosure, each side generally being used for opposite directions of transmission. However, both sides of the repeater can also be made to operate in the same direction. Because of the configuration of the

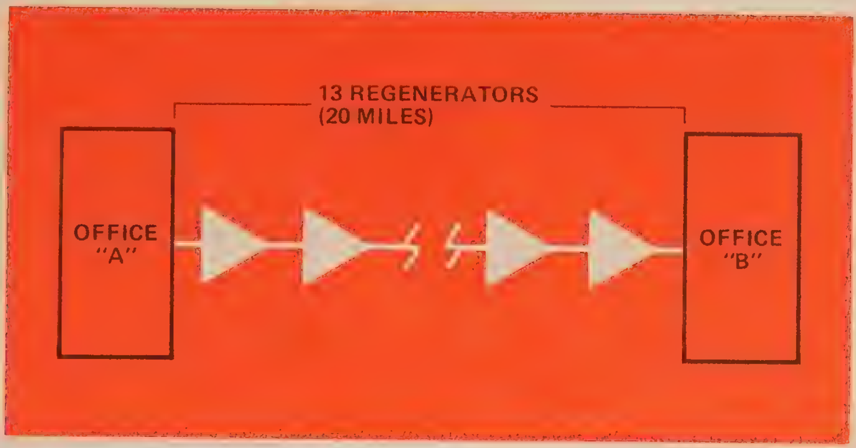


Figure 1. The field study was conducted between two terminals 20 miles apart.

splicing plan, a 500-mile, two-way system (325 repeaters), or a 1000-mile, one-way system (650 regenerators), could be assembled. This permitted measurement of the accumulation of phase jitter as the length of the line was increased.

Test Signal

All the testing was done at office "A", with alternate pairs looped back at office "B". This provided 40-mile increments of 26 regenerators each. The test setup for the facility is shown in Figure 2. Phase jitter was measured from oscilloscope patterns, with a special signal generator providing the driving signal so that jitter caused by pattern changes could be observed. The most prevalent kind of phase distortion in a PCM line is caused by low-frequency phase jitter, which arises when pulse-pattern changes take place.

The special signal generator was constructed to give two pulse-shift patterns. Each output consisted of a 1.544-Mb/s bit stream which had two fixed patterns, with a shift between the patterns at a regular rate. These pulse patterns were as follows (see Figure 3):

1. from one pulse out of eight time slots (1/8) to two pulses out of eight time slots (2/8), shifted at a 2-kHz rate.
2. from one pulse out of eight time slots (1/8) to one pulse out of four time slots (1/4), shifted at a 2 kHz rate.

These pulse-shift patterns were selected to give representative worst case jitter. With a fixed pattern through all tandem regenerators, each regenerator clock assumes a fixed phase. When the pattern shifts to another fixed pattern, each regenerator clock shifts phase in relation to its previous setting. The phase shift propagates down the line, and is the same for each regenerator.

Study Results

The repeatered line was looped back at office "B" as shown in Figure 2, to allow jitter to be measured in 26-regenerator increments (40 miles), from 26 regenerators to 650 regenerators (1000 miles). The output of the repeatered line at office "A" was displayed on an oscilloscope, with external synchronization supplied by the clock of the signal generator at the input to the line. Low-frequency phase jitter in nanoseconds could be read from the scope as the line length was increased. This would show jitter accumulation. Figure 4 shows a photograph taken from the oscilloscope face which graphically displays the offset of the pulses as the pattern shift takes place (after 26 and 182 regenerators). Figure 5 shows the cumulative jitter as

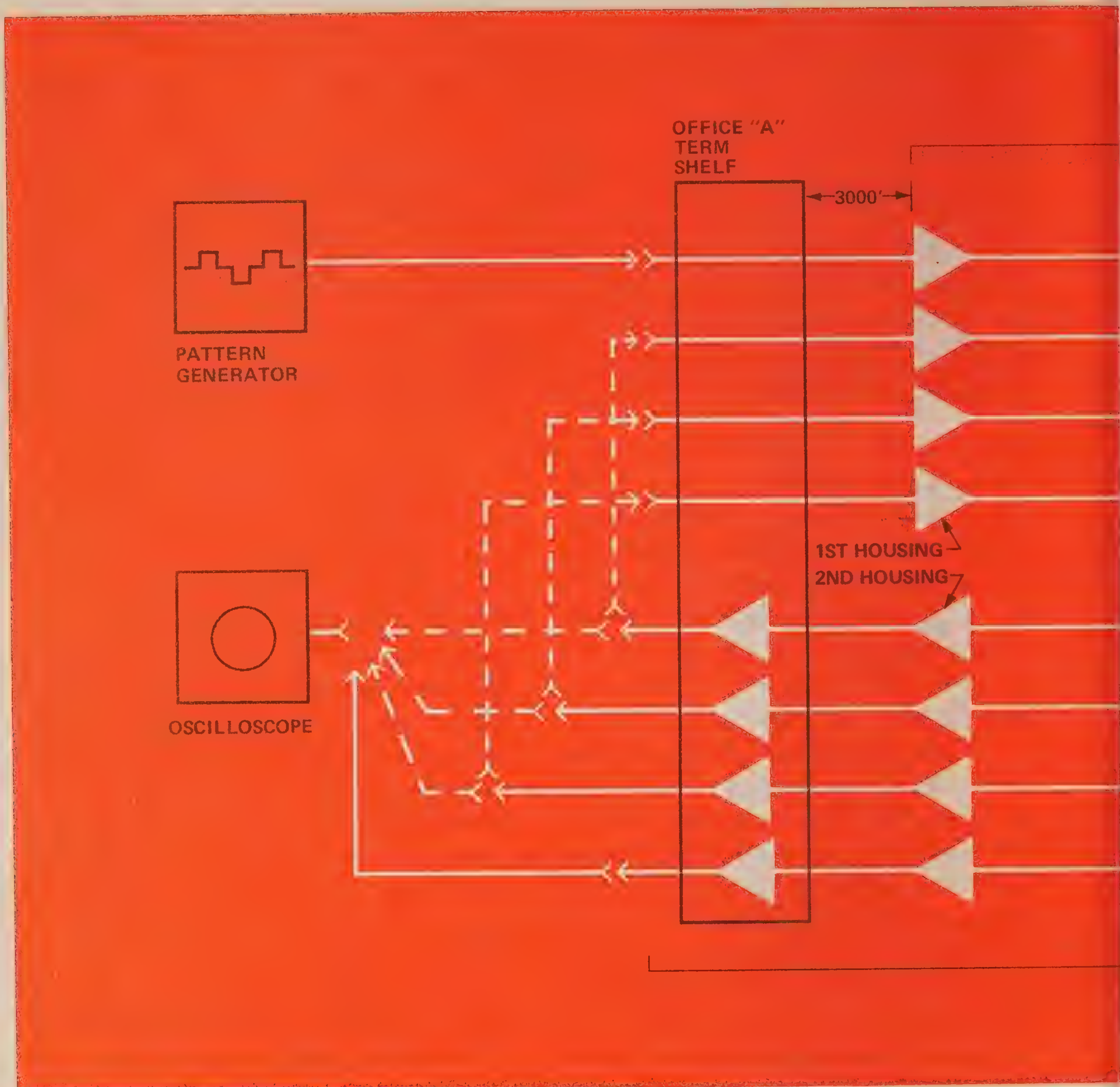
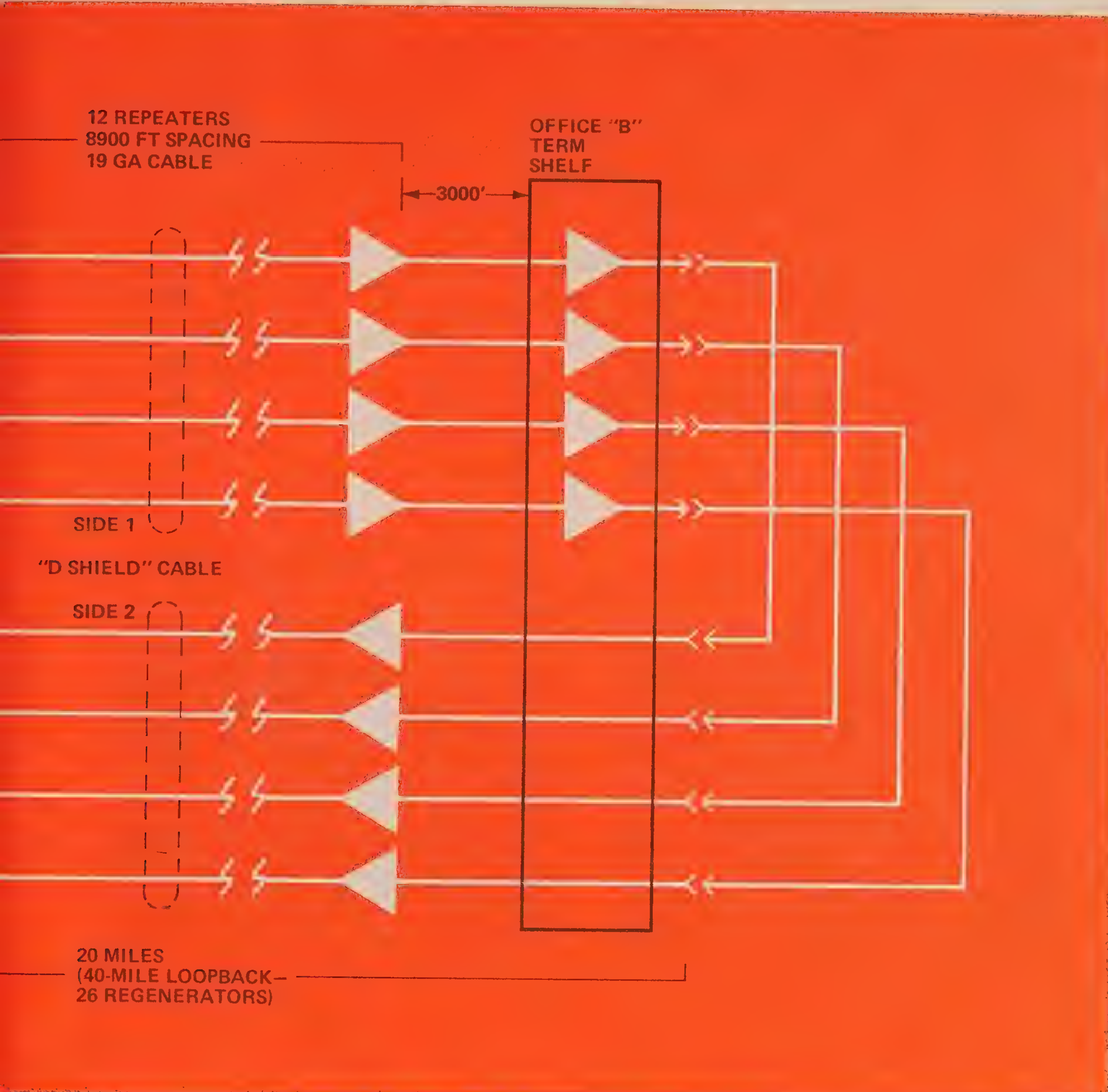


Figure 2. The PCM line was looped back at office "B" to give 40-mile sections in 26-regenerator increments.

the line length increased. Theoretically, low-frequency jitter should be directly additive, with each regenerator adding an equal amount. Calculations made prior to the field study indicated that the worst-case jitter added per GTE Lenkurt 9101C repeater should be about 2-1/2 nanoseconds peak-to-peak, and should add linearly regardless of the number of regenerators in tandem. The measured results in Figure 5 substantiate the expected linearity for the number of tandem regenerators that were measured.

Another cause of phase jitter is the coupling from one side of a repeater (with one pattern) to the other side (with another pattern). A study was made over the 500-mile, two-way system to determine if this effect was significant in the 9101C regenerators. The results of this study showed that any contribution due to this source could be neglected.



Effect on PCM Terminals of Long Repeatered Lines

Theoretically, a PCM terminal should operate without degradation over any length of repeatered line because of the regeneration capability of the line. To test this theory, two 24-channel GTE Lenkurt 9002A terminals were connected end-to-end with a 500-mile repeatered line between them. Idle noise and quantizing distortion were selected as the best figures of merit to indicate whether the terminal was operating properly. Idle noise was measured on all 24 channels at each terminal of the 500-mile system. Compared to a channel bank specification of 23 dBrnc0 maximum, all channels were within 19 to 23 dBrnc0 (averaging approximately 20 dBrnc0). Quantizing distortion was also measured on one channel at each terminal. Quantizing distortion, which is the single most sensitive measurement in a PCM system, proved to be essentially the same as on a looped-back channel bank with no repeatered line connected.

The study conducted showed that (at least for the GTE Lenkurt 9101C

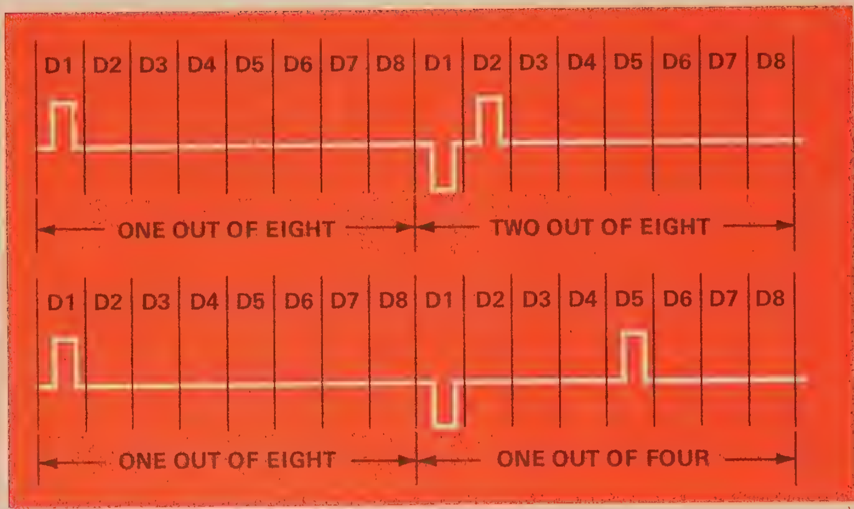


Figure 3. The special signal generator used in the field study had two outputs. One was a pattern shift from one pulse out of eight to two pulses out of eight; the other was from one pulse out of eight to one pulse out of four.

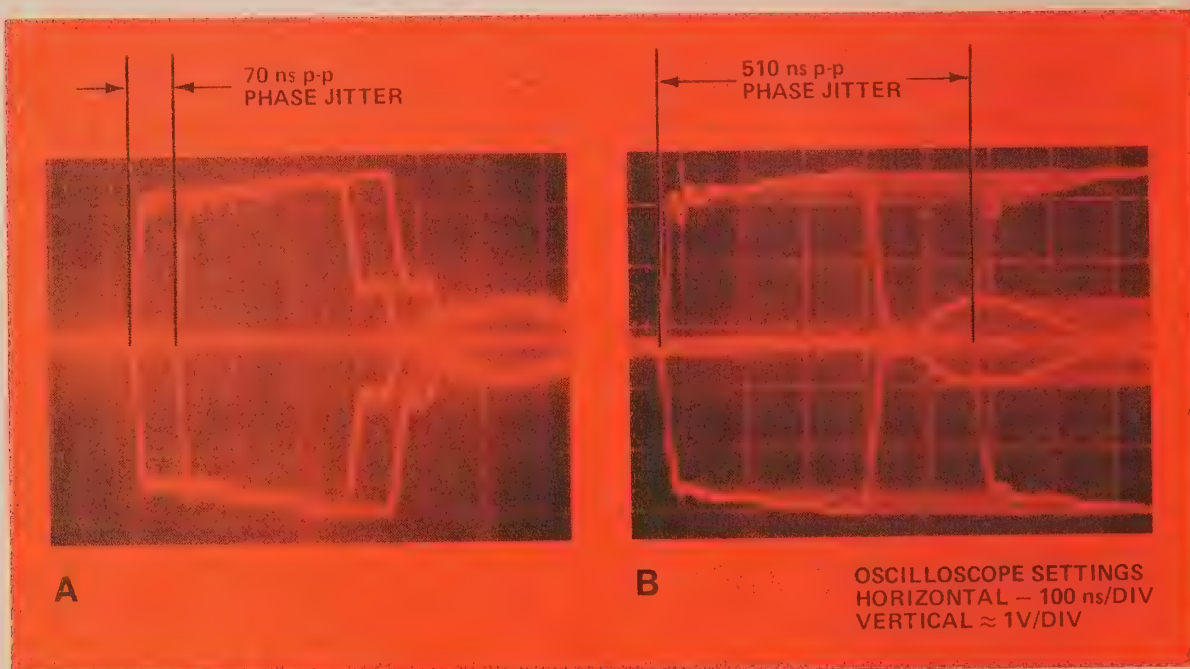


Figure 4. Using the one-out-of-eight to one-out-of-four pattern shift, the cumulative effect of phase jitter can be seen by these two oscilloscope photographs, A is with 26 regenerators and B is with 182 regenerators.

system) channel bank performance is not significantly affected by the number of regenerators placed in tandem, in spite of systematic jitter build up. If indeed there is a practical limit placed on the length of a T1 line between channel banks, it will most likely be imposed by administrative restrictions, not degradation.

The study also verified the calculations that GTE Lenkurt 9101C repeaters contribute some 2-1/2 nanoseconds of phase jitter per regenerator and that it accumulates linearly with the number of tandem repeaters. From this, it has been determined that no more than about 200 repeaters should be used in a single system, if it is anticipated that future multiplexing with multiplexers like the GTE Lenkurt 9120A (3 Mb/s on cable or radio, 6 Mb/s on cable) will take place over T1C (48 channels, 3 Mb/s) or T2 (96 channels, 6 Mb/s) lines.

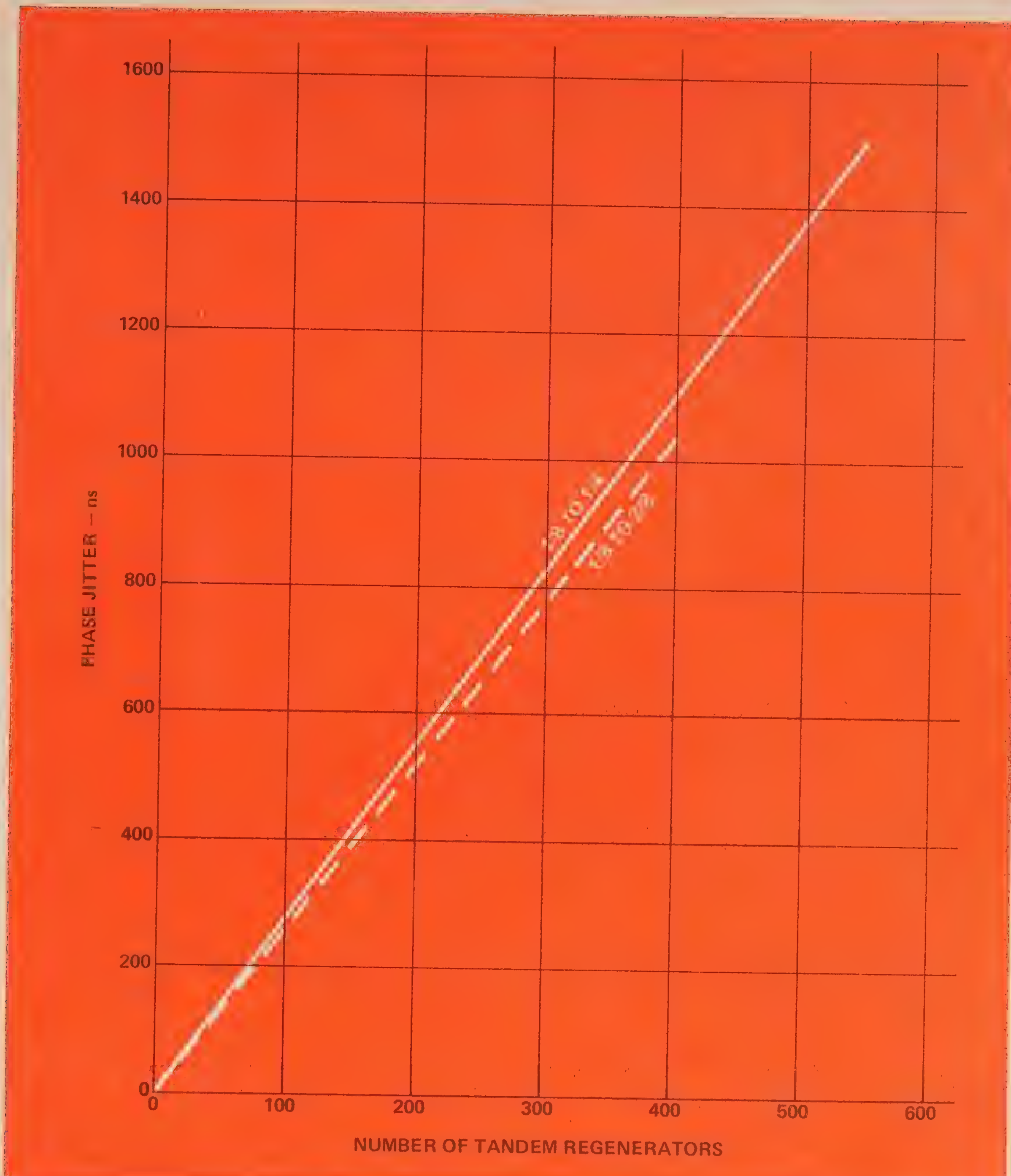


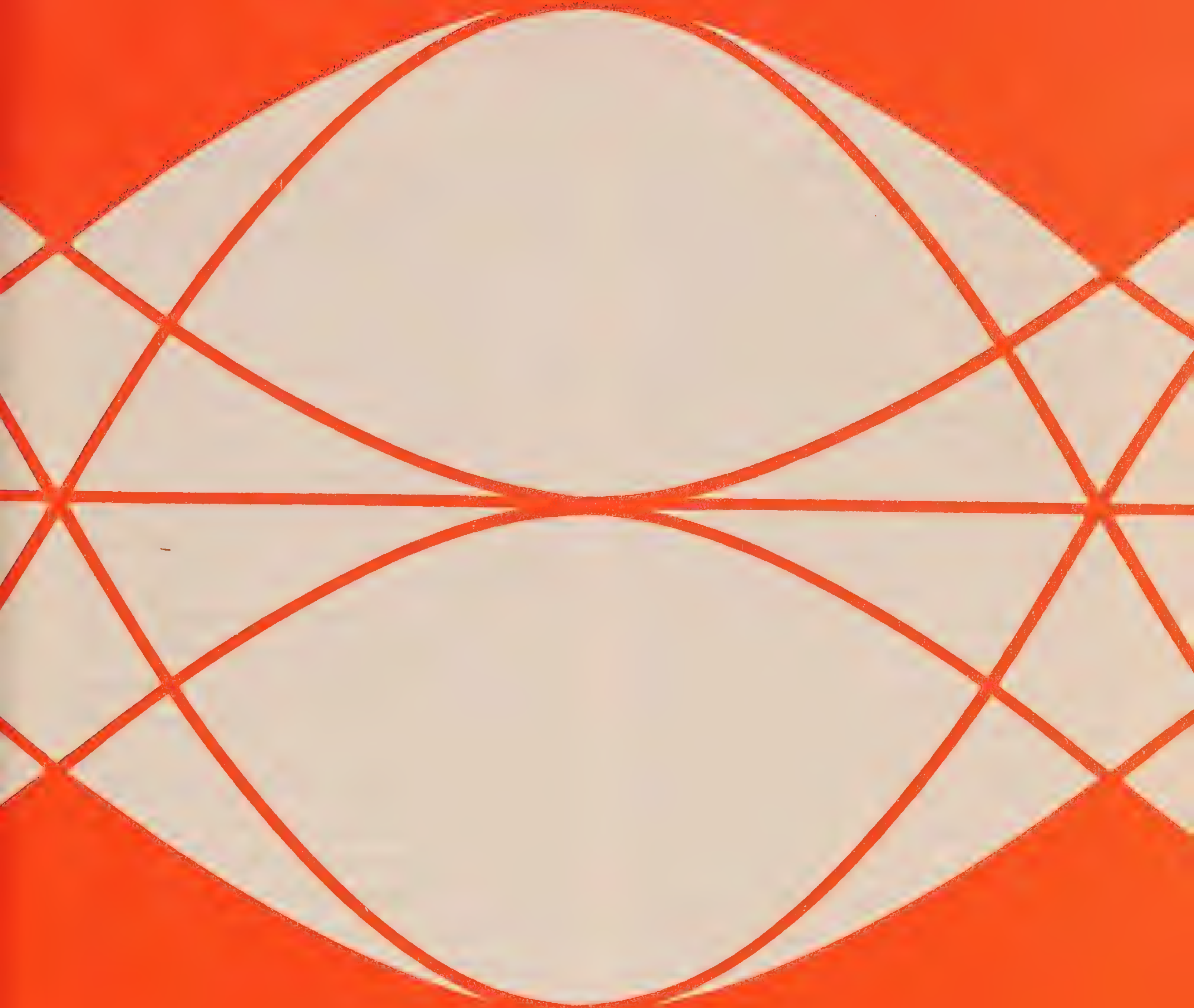
Figure 5. The field study proved that phase jitter adds linearly for both pulse-shift patterns.

BIBLIOGRAPHY

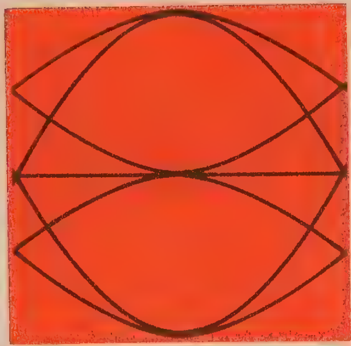
1. Crawforth, L.D. and J.D. Olson. "Maximum T1 Repeatered Line Length Considerations," *International Conference on Communications*, Vol. II, (June 1973), 32-28–32-33.

2. "9101C Repeatered Line Measurements," San Carlos, California: GTE Lenkurt, 1973.

GTE LENKURT
DEMODULATOR
JUNE 1973



PCM Cable Considerations



Pulse Code Modulation is finding greater usage in today's modern communications systems, and promises to be one of the major forms of telecommunications in the future. Yet, the equipment that makes PCM a reality, is greatly dependent on the cable used for transmission.

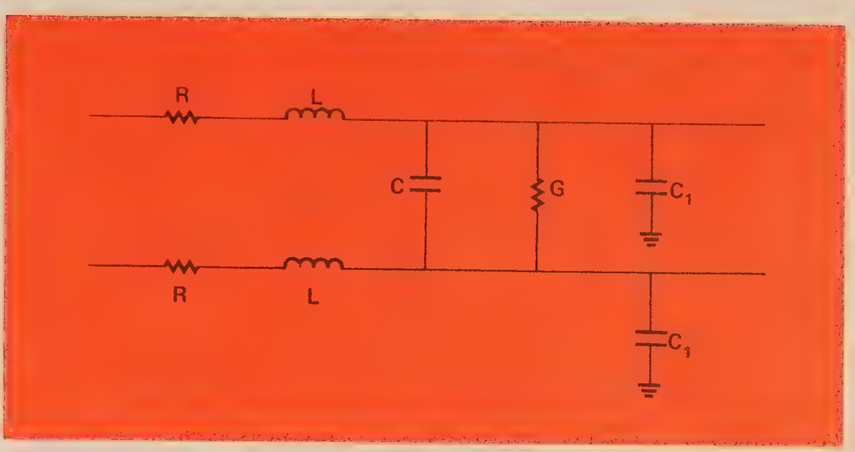
Cable pairs that are to be used for PCM, should be checked for suitability prior to installation of the carrier and repeater equipment. The usual method of testing cables for PCM is with a frequency selective VTVM (vacuum tube voltmeter) and a signal generator or oscillator tuned to 772 kHz, which is the equivalent fundamental frequency of a T1-type PCM pulse train. With the signal generator transmitting at 772 kHz at one end of the cable pair, the amount of signal attenuation can be ascertained by the VTVM at the other end of the line. This method has considerable limitations in that it tests cable performance at only one frequency – 772 kHz. Theoretically, a cable pair *could* be suitably checked with a signal generator if all frequencies from zero to approximately 2.5 MHz were transmitted individually down the cable pair, but this would indeed be a tedious procedure. The cable is also checked for correct loop resistance, insulation, capacitance and dc continuity. (See previous *Demodulator* issues for descriptions of general PCM operation.)

Differences in Transmission

It is impractical to assume that a cable that is known to operate satisfactorily in the voice frequency band will necessarily accept PCM signals. The fault in this assumption is that there are more critical transmission requirements for the high bit rates associated with PCM. A cable pair acts as a low-pass filter to transmitted signals. It has a tendency to pass the lower frequencies quite easily, but rejects the higher-frequency components. In rough figures, the loss per mile at voice frequencies (0 – 4 kHz) is about 1 dB, but about 31 dB per mile at 772 kHz.

Although simple in appearance, a cable pair has complex electrical properties which must be taken into consideration when designing a transmission system, since they determine the transmission characteristics of the cable pair. Figure 1 shows a simplified equivalent circuit of a cable pair. The series resistance (R) is the simple ohmic resistance of the conductors. The series inductance (L) is the self inductance of each conductor, plus the mutual inductance between the individual conductors. Shunt conductance (G) is the total conductance of the current leakage paths between the conductors, and shunt capacitance (C) is the electrical capacitance between conductors. C_1 is the capacitance between the conductors and ground. These parameters define the attenuation characteristics of the cable, which can be represented as an attenuation-

Figure 1. The electrical properties which determine a cable pair's suitability for PCM are distributed along the entire length of the cable.



versus-frequency curve. The particular values of these parameters depend on the physical configuration of the cable, the material of which it is constructed, the frequencies involved, and the ambient temperature. The simplified equivalent circuit shown in Figure 1 shows the cable parameters as lumped constants, but in reality, they exist uniformly along its length, and are considered to be distributed rather than lumped. The inductance in a cable tends to act as an open circuit at high frequencies, and the capacitance approaches a short circuit at high frequencies. As frequency increases, the more loss there will be in the cable.

Because of the frequency difference between the voice band and the spectrum used for transmission of PCM pulses, acceptable voice frequency pairs may not be good PCM pairs. Voice usage only requires that a cable pair have consistent properties out to about 4 kHz. PCM requires that the cable perform as predicted out to about 2.5 MHz. Added to this requirement of a very wide bandwidth are other potential problems such as crosstalk and impulse noise which become more critical at the higher frequencies. It is important for a user who is contemplating a PCM-system installation to know if his existing cable meets the necessary requirements. A simple solution to accurate determination of cable suitability for PCM is to simulate an actual PCM pulse train on the line before installation of carrier equipment. This can be done quickly and inexpensively by using one of the PCM cable test sets recently introduced in the telecommunications industry.

PCM Cable Test Sets

One such test set is the GTE Lenkurt 91100 PCM Cable Test Set, which consists of a two-section unit that can transmit or receive a bipolar, T1-type PCM pulse train at 1.544 megabits per second. Two test sets are necessary for PCM cable section testing. The pulse train from the test set is "pseudo random." It is random in that any combination of ones and zeros may appear on the line, but pseudo random in that the pattern repeats itself after a set number of transmitted bits. The receiving section of the set evaluates the incoming pulses, and displays on a meter and dial, the cable's suitability to accept PCM signals. Figure 2 shows a block diagram of the PCM cable test set.

In the laboratory, it is possible to see the way in which the test set evaluates a cable pair by observing the eye pattern produced when an oscillo-

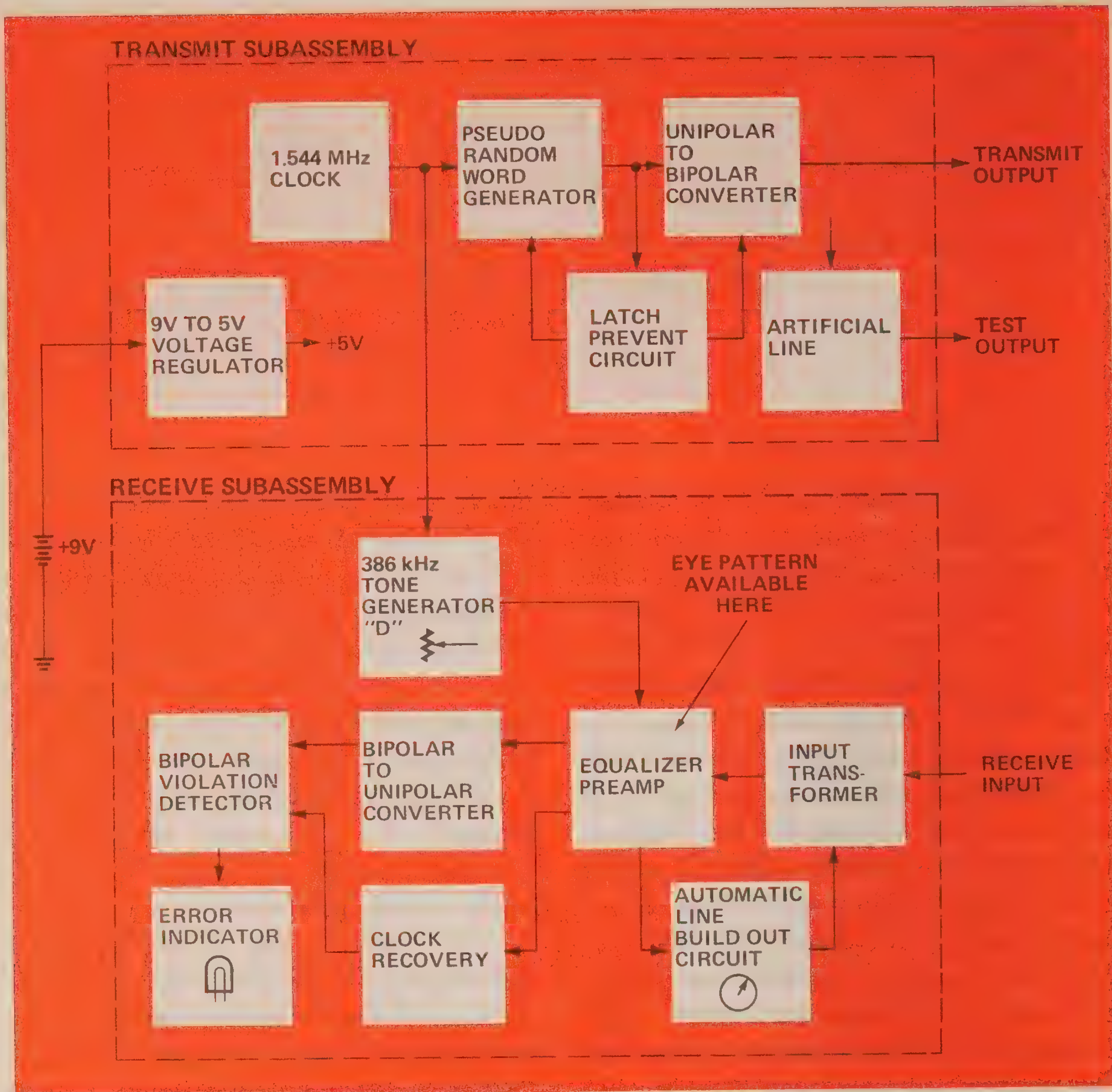


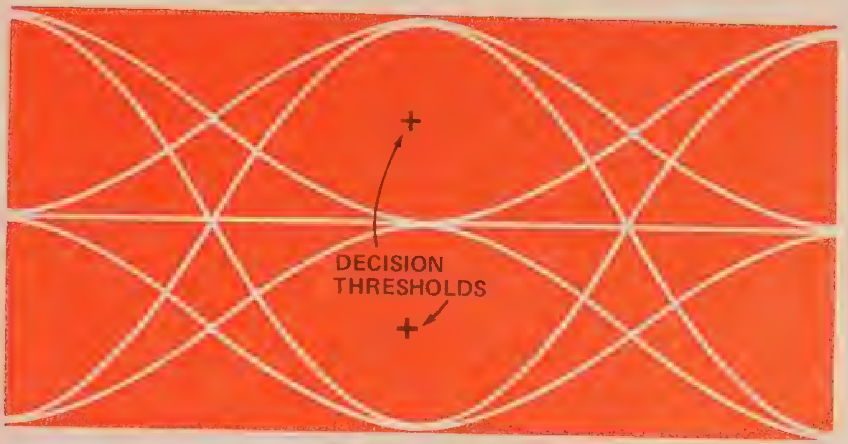
Figure 2. Block diagram of a PCM cable test set.

scope is connected to the preamplifier of the set. The oscilloscope displays all the incoming bipolar pulses superimposed on one another. Figure 3 shows an example of an ideal eye pattern. The threshold detector in the test set is set at the center of the eye, where there is maximum probability of detecting a one or a zero. Any signal that goes above the threshold is a one, and any signal that falls below, is a zero. The eye pattern is a composite of all the incoming pulses and the distortion associated with them. It makes visible to the observer, the condition of the incoming digital bit stream by displaying an open, partially-open, or closed eye. Figure 4 gives an example of an open eye and a partially-open eye under actual transmission conditions. The amount of closure is determined by the amount of distortion in the cable.

Noise Margin

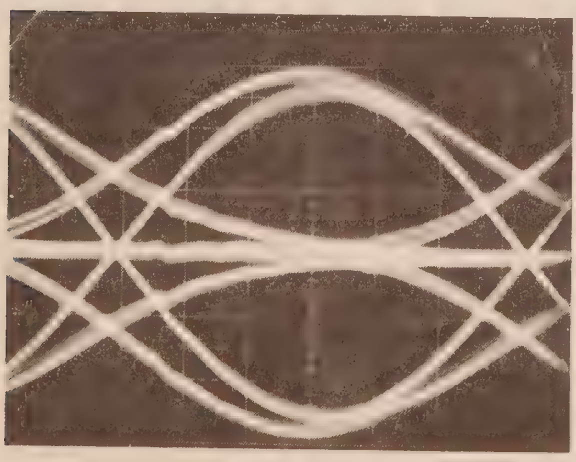
The noise margin of the cable determines how much more noise a cable can tolerate before it is no longer suitable for PCM transmission. Figure 5 shows how the test set may be

Figure 3. An ideal eye pattern which would be observed with a perfectly noise-free signal.

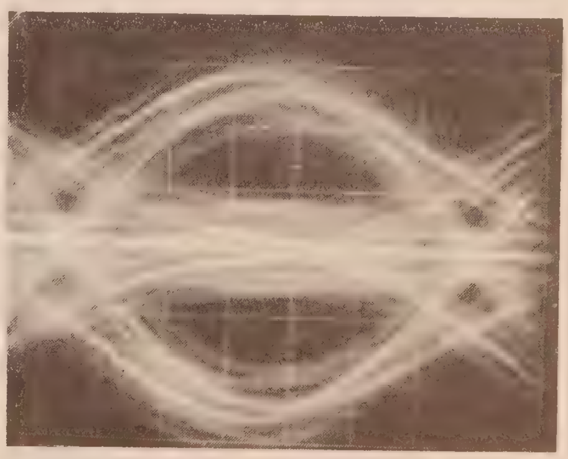


utilized to check cable sections between repeaters and complete span lines between terminals. If the test set at A transmits a pseudo-random pulse train which is received at test set B, the meter at B records the amount of attenuation in dB that the signal has undergone in traveling from A to B. If, for example, a 6,000-foot section of 22-gauge, .083-μf/mile cable were under test, and the attenuation was 27 dB, this would be an acceptable loss for that length of cable. However, a much higher dB reading would indicate a bad cable or the presence of some obstruction such as a loading coil or building-out network. The use of existing cable pairs for PCM or any other type of carrier service requires that they be free of loading coils, building-out networks, crosses, splits, high-resistance splices, grounds, moisture, and bridged taps.

If the amount of attenuation is acceptable, the operator at the receive end begins to apply an interfering tone to the incoming signal, by turning a D-factor potentiometer. The D-factor corresponds to a mathematical relationship between the ideal eye and the eye of the cable under test. The D-factor potentiometer indicates the amount of *degradation* imposed on the incoming signal by the test set. The amplitude of the interfering tone is increased until the test set begins to make errors. Visually, this would create a closed eye pattern, and would



OPEN EYE



PARTIALLY-OPEN EYE

Figure 4. The greater the eye opening, the better the possibility that the cable is suitable for PCM.

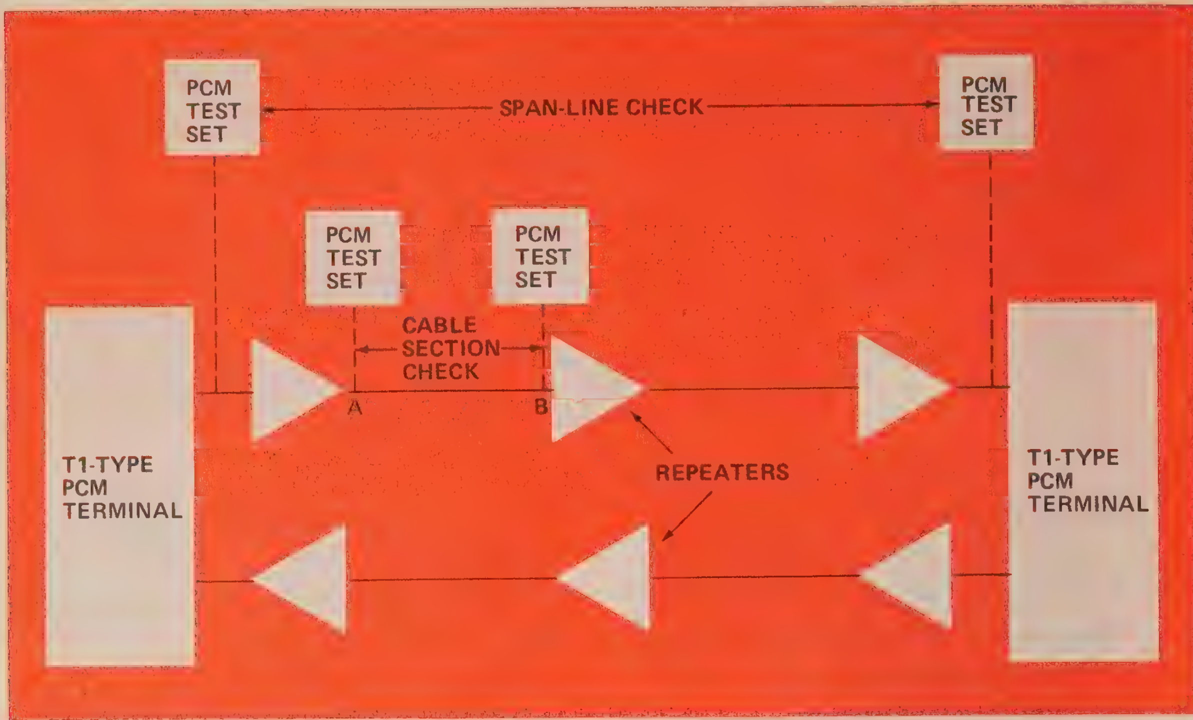


Figure 5. A PCM test set can be used to check cable pairs between repeaters, and entire repeatered lines.

correspond to a situation where the threshold detector of a repeater, or in this case, of the test set, could no longer distinguish between a one and a zero. The reading on the D-factor dial is numbered from 0 to 1.0 in 0.1 steps, with the lower numbers indicating the best conditions. Figure 6 shows the D-factor readings and the corresponding loss in margin-to-noise (system degradation). In practice, desired D-factors for normal cable sections will be between 0.1 and 0.3. Cables with D-factors between 0.3 and 0.5 are still acceptable, but as shown in Figure 6, the margin is reduced 6 dB at 0.5 compared with 3 dB at 0.3. Cable sections with D-factors over 0.5 are not suitable for PCM transmission.

Tests In Both Directions

It is important to test the cable in both directions, since a faulty cable may check good in one direction, but bad in the other. If, for example, there is water in the cable between the two test sets, it will appear as capacitance at a localized point. In a 6,000-foot section of cable, if there is water at 1,000 feet from the transmitting test set at point A, most of the energy will be reflected and the receiving test set at point B will detect a fault and give a high D-factor. However, when point B transmits to point A, since the signal has already traveled 5,000 feet before it hits the water, the reflection will be less, and the reading on the D-factor scale may show the cable as acceptable. So, when the reading in one direction is much better than in the other, it is an indication of a fault in the line.

An important application of the test set is in checking the performance of the complete span line, including line and office repeaters. This test will reveal marginally-operating repeaters, since a span line that can handle the pseudo-random signal generated by the

D-FACTOR	LOSS IN MARGIN, dB
0.1	1
0.2	2
0.3	3
0.4	4.5
0.5	6
0.6	8
0.7	10.5
0.8	14
0.9	20
1.0	INFINITE

Figure 6. The lower the D-factor, the better the suitability of a cable pair for PCM.

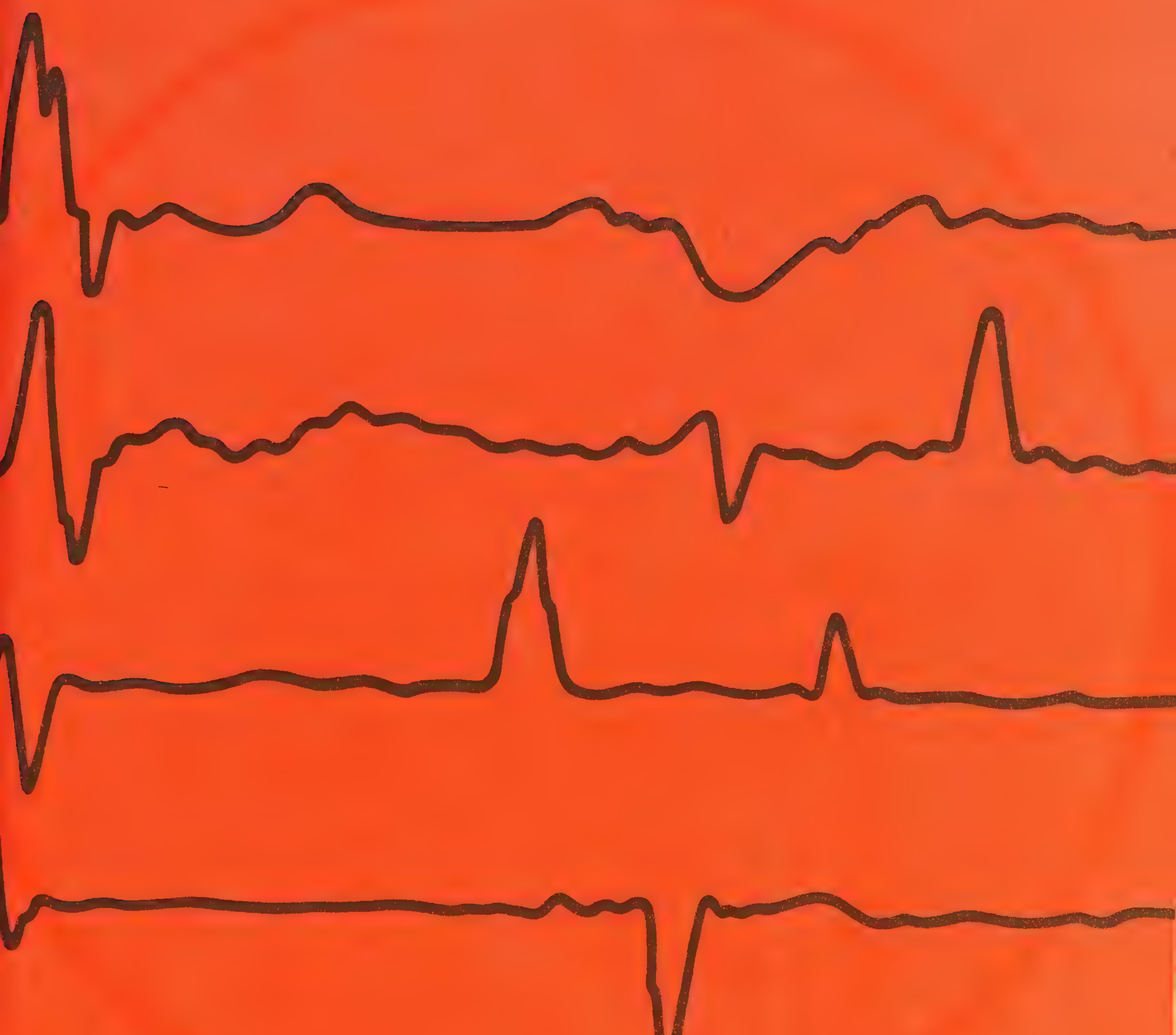
test set will perform well with the signal of an actual PCM system.

Additional Systems

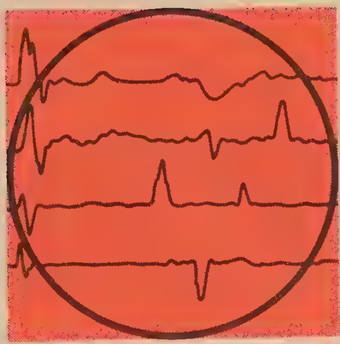
When adding a PCM system to a route that has been designed for N number of systems, but due to growth, requires another system, it must be assured that the new system will not cause failures in the existing systems. The cable pairs for the new system should be tested by D-factor measurements in each repeater section. At the same time, any adverse effect that the new system might have on working systems in the same cable can be observed. While one test set is transmitting in the forward direction, errors may be checked on working systems at the receive terminal using a test set or built-in error detectors on the office terminating equipment. If errors do not increase, there is added confidence that the new system will not disturb systems already in service. This is especially so since the sustained pseudo-random output of the test set signal will produce more crosstalk than the normal signal of a PCM system.

The usual dc methods of testing cable such as checking loop resistance, insulation of cable, capacitance of cable, and dc tests for continuity still provide valuable information on the condition of the cable, but the use of cable test sets that send a true PCM signal down the line insures that the cable will stand up under actual PCM-transmission conditions, and evaluates how a cable with existing PCM systems will bear additional systems.

GTE LENKURT
DEMODULATOR
DECEMBER 1971



CABLE TESTS AND MEASUREMENTS FOR PCM



For the user of communications equipment who is contemplating connecting new equipment to cable that has already had considerable service, it is wise to be sure that existing lines are in acceptable condition.

New equipment has on many occasions been connected to older cable resulting in an unsatisfactory system. In a situation such as this, if only part of the total number of channels function properly – what is at fault, the cable or the equipment? Because it is difficult to reliably estimate the overall condition of older cable without testing, this question can only be answered after a complete series of tests and measurements have been performed on the cable.

Because of its period of usage, deterioration of older cable may result in such faults as shorted cable pairs or moisture within the sheath. To detect faults such as these, measurement of each cable pair is still the ideal method of operation. Although the measurements discussed here tend to favor testing for PCM, they are suited for all types of carrier operation.

The use of cable pairs for PCM or any other type of carrier service requires that they be free of loading coils, building-out networks, crosses, splits, high resistance splices, grounds, moisture, bridged taps, and cable terminals bridged across pairs (when subscriber cable is used). Should a bridged tap, loading coil, or building-out network scheduled for removal be overlooked when reading a cable location map, subsequent visual observations may completely fail to detect it. For this reason, pulse reflection tests using a radar test set are recommended when the presence of any of these external additions is suspected. Additional tests which help to evaluate the overall condition of a cable include:

(1) DC loop resistance and conductor resistance balance test using a Wheatstone bridge.

(2) Insulation resistance measurement using a megger test set.

(3) Frequency response test to 1 MHz using an oscillator and frequency selective voltmeter.

Radar Cable Test Set

The radar cable test set is an invaluable tool in detection of cable discrepancies and attached networks. The test set, as its name implies, operates on the simple principle of pulse reflection. The nature of the return pulse reveals the type of discontinuity present in a cable. For example, a positive (upward) deflection on the test set oscilloscope trace indicates an open circuit or high impedance mismatch while a negative deflection indicates a short circuit or low impedance mismatch. A shorted pair returns a strong negative pip at the point of short circuit since almost all energy is returned from the fault. A good cable pair will show either no reflection or an open (positive pulse) providing the end of the cable is within the range of the radar test set. In addition to

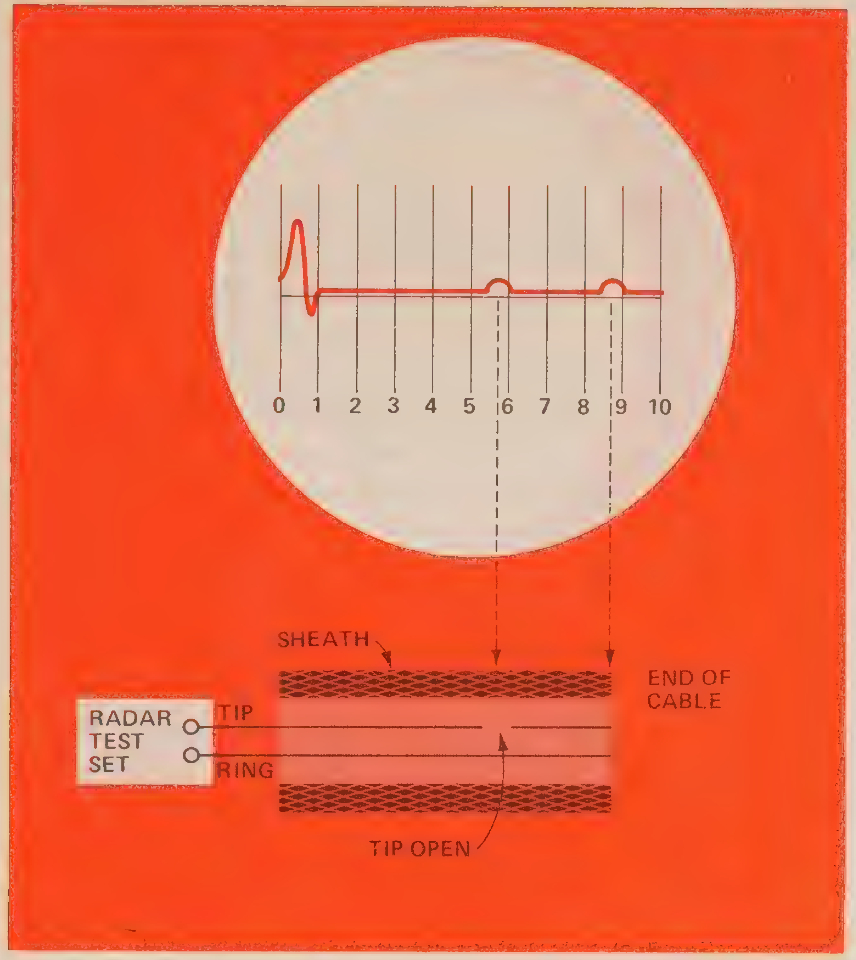


Figure 1A. A pair with one conductor open will show a positive return pulse.

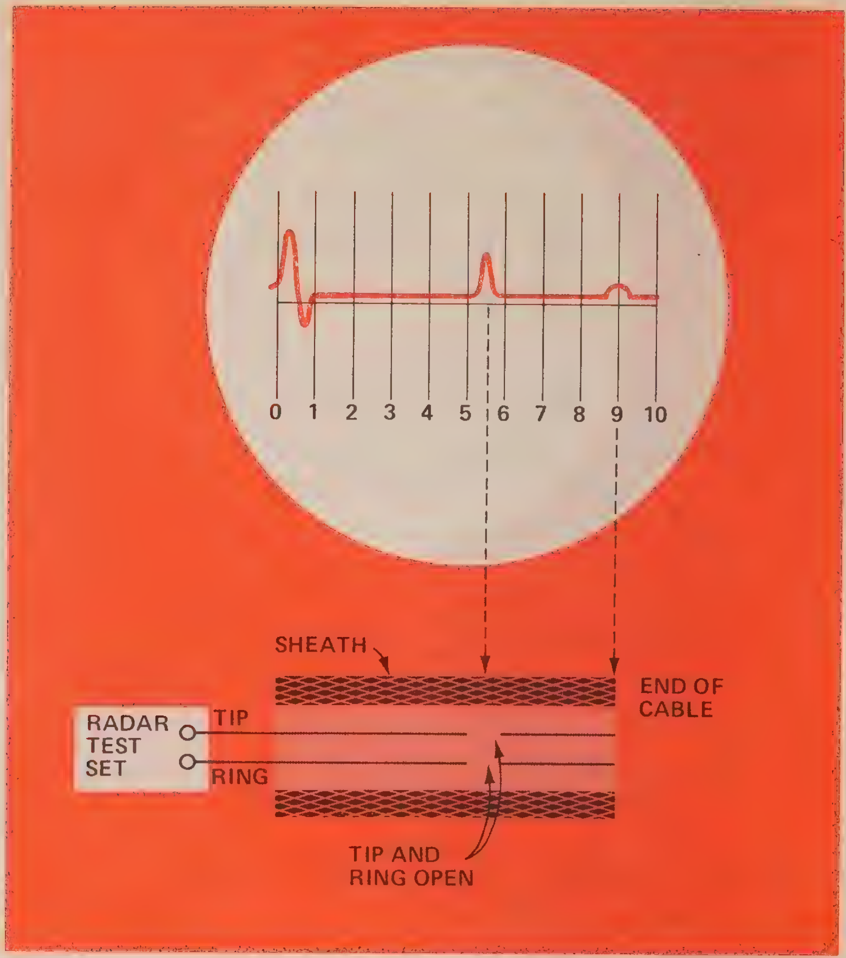
revealing the nature of the discrepancy in a cable, the radar test set can also show the actual distance to the point of fault.

Although the operation of a radar test set is relatively simple, some experience is necessary on the part of the operator in the interpretation of the return signals. For example, a negative return pulse displayed on the oscilloscope of the radar test set for two separate cable pairs does not necessarily mean that the cable pairs have similar discrepancies. The method of detection in this case would be to notice the amplitude of the return pulse since different amplitudes indicate different cable faults.

Open Conductors

An open in either conductor will result in a positive pulse on the oscilloscope screen although of lesser amplitude than if both conductors were open. Figures 1A and 1B show the oscilloscope display as it would appear with one and two conductors open. Also shown are the corresponding cable pair and sheath with the appropriate conductor designations. The reflection from the end of a cable may sometimes be observed in spite of an open in one of the conductors, but this depends upon the closeness of the fault to the end of the cable and upon the size of conductor used. Telephone communications lines generally use cable which ranges from 16 to 26 gauge; the attenuation of the cable increases in direct proportion to the gauge number. The return pulses in 26-gauge cable are the weakest since this gauge has the greatest attenuation due to its higher resistance. Testing

Figure 1B. When both conductors are open the return pulse will be of greater amplitude than that for one conductor open.



a 26-gauge conductor, the maximum fault location distance is about 4000 feet. For greater distances (8000 to 10,000 feet) a special pulse amplifier must be used to see the return signal.

Historically, the tip (T) and ring (R) designations were so called because they corresponded to the contacting part at the tip and ring of the phone plug used to make circuit connections in a manual switchboard. Today, the designations tip (T) and ring (R) are used to identify the two conductors of a cable pair. The sleeve of the plug is used for certain control functions, not directly associated with the cable pair.

Loading Coils and Building-out Networks

The function of a loading coil is to increase line inductance and thereby improve transmission characteristics. As shown in Figure 2, the tip and ring conductors connect to separate windings on the donut-shaped core. The two coils are wound in a direction that produces an aiding magnetic field. A cable which runs between two points is engineered according to a predetermined loading plan. For example, an H88 loading plan requires 88-millihenry coils to be placed at 6,000-foot intervals along the cable.

If for some reason, be it geographical obstruction or inconvenient location, a loading coil cannot be placed in the designated area, a building-out network is used to artificially make up the required distance (6000 feet in the case of H88 loading). The building-out network consists of resistors and capacitors connected in such a way that electrically, the distance between load-

ing coils conforms to the required loading plan. That is, the building-out network contains the resistance and capacity inherent in the length of cable necessary to conform to the loading plan. Loading coils and building-out networks must be removed from the cable before it can be correctly evaluated for use.

Open Sheath Detection

A high-frequency pulse such as is emitted from the radar test set will usually be prevented from passing through a loading coil by the choking effect of inductance on high frequencies. However, using a certain hookup procedure, the radar test set may be made to measure beyond loading coil points.

Individually, the tip and ring windings of the loading coil will have a nullifying effect on a high frequency pulse, but when the pair conductors are connected together, this effect is counteracted to some extent.

To detect an open sheath, one or more cable pairs which have been tested and found to be serviceable must be connected in parallel to one terminal of the radar test set; the other terminal is connected to the cable sheath. In this way the effect of the loading coils is cancelled and the test pulses can pass beyond the attenuating inductance. Figure 3 shows the radar test set hookup and corresponding oscilloscope display for open sheath detection.

Grounded Sheath On Buried Cable

When cable is buried by means of a cable plow, the surrounding earth will remain loose, thus creating a void around the cable. With the passage of time a gradual repacking takes place and the void is eliminated. Should a grounded sheath exist during a void environment, detection of the fault is possible provided the void is filled with water such as occurs in an area with a high water table or after a rainfall which has heavily wetted the ground.

Plowed cable normally disturbs the surrounding earth to such an extent that the sheath-to-ground capacitance is non-uniform. This causes sufficient variation in capacitance to make propagation of the radar pulses difficult. However, when the void is filled with

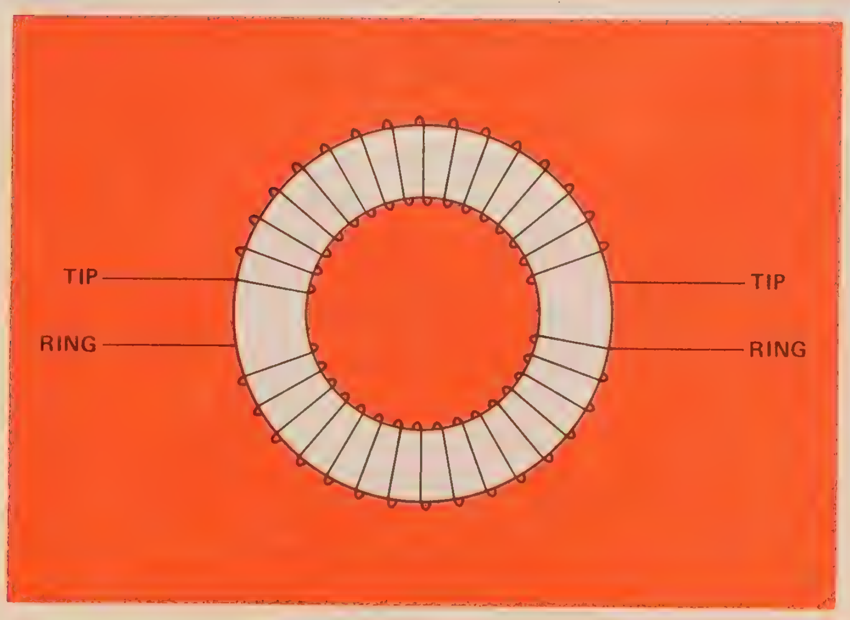
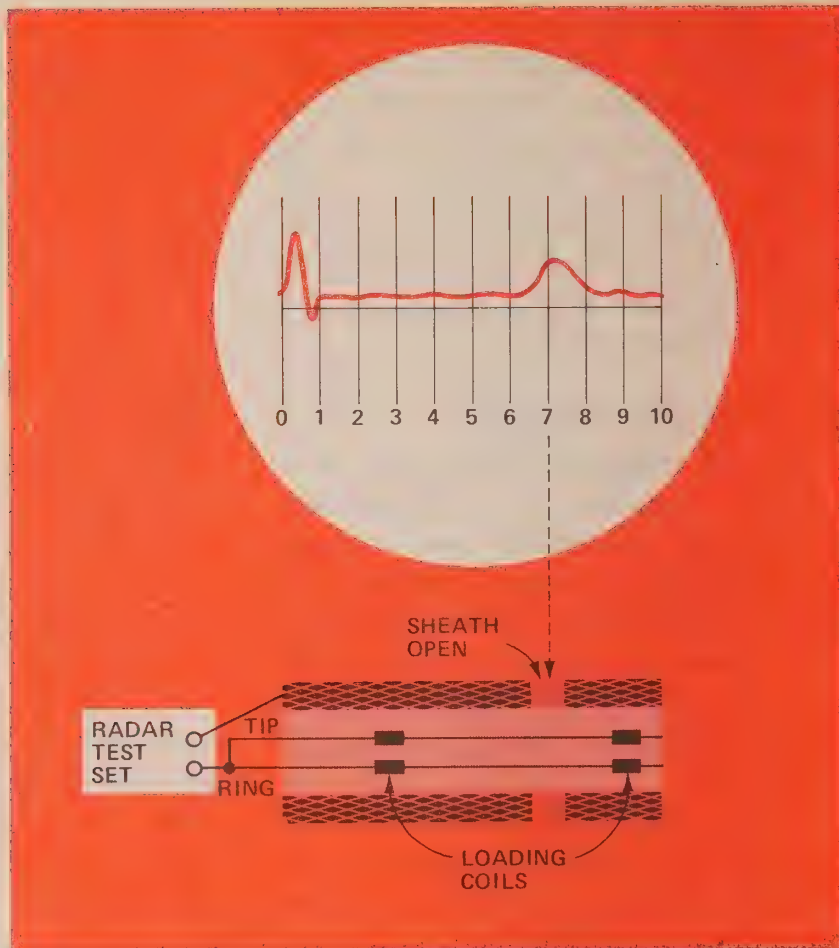


Figure 2. Loading coil with corresponding tip and ring connections.

Figure 3. Display and hookup for open sheath detection.



water, or earth conductivity is low (40 to 50 ohms), the capacitance tends to become more constant, and a pulse may be returned with sufficient strength to be visible. Figure 4 shows the grounded sheath oscilloscope display with the corresponding test set connections.

Cable Splices

Due to the change in relationship between cable pairs and the sheath, a splice represents a decrease in capacitance, and will appear as a small rounded positive pip (radar return pulse) followed by a small negative pip on the oscilloscope trace. At times, the negative pip may not appear. A known splice may be used as a reference point for calibration when performing measurements in a cable of unknown dielectric constant.

Crosses

Metallic crosses (short circuits) between pairs present somewhat the same indication as a short to the sheath. Since a metallic cross is usually a solid connection it may seem that the easiest way to find the other pair or conductor involved is by performing a standard battery and earphone test. However, since the radar test set indicates the distance to the trouble, once the sheath is opened the trouble can be found visually.

Crosses due to moisture in the cable appear as an extremely noisy trace with a vertical displacement of the entire trace throughout the wet section, as shown in Figure 5. Any reference points, such as splices within the wet area will appear farther apart than normal. Ranges in the wet area will appear approximately 60% greater

Figure 4. Grounded sheath on buried cable.

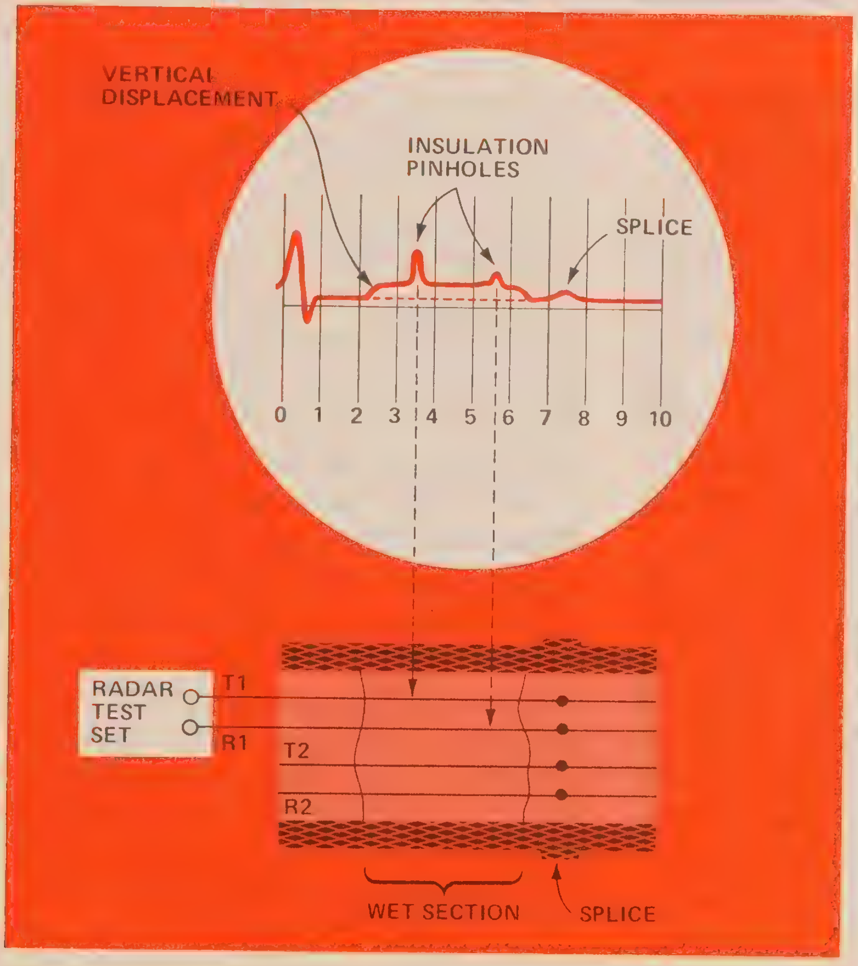
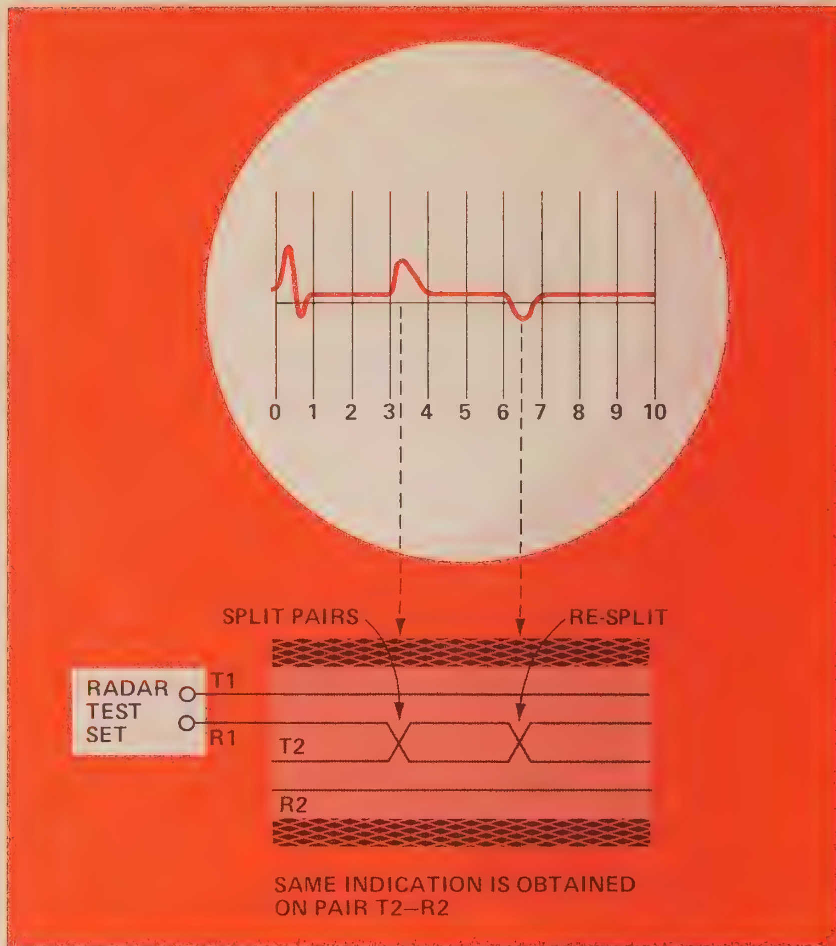


Figure 5. Crosses due to moisture appear as an extremely noisy trace with a vertical shift.

Figure 6. A split pair on a cable causes a decrease in capacitance and a positive pulse on the scope.



than the reference distance if the fixed Polyethelyne Dielectric setting on the radar test set is being used. This setting takes into consideration only dry PIC cable, the PVC (propagation velocity constant) of which is 0.667. This velocity constant means that high-frequency signals in the cable will travel only 66.7% as fast as they would travel in free space. The retardation effect caused by the higher dielectric constant of water decreases the propagation velocity of the pulse, yielding a slower traverse of the pulse through the wet section. To accurately measure the distance within the wet area, the Cable Dielectric Switch on the radar test set may be manually adjusted until a known reference point such as a splice is correctly positioned. The correct distance may now be read in spite of a change in PVC. The propagation velocity constant of wet cable is approximately 0.400.

Split Pairs

A split occurs at a splice when the tip or ring of one cable pair is accidentally spliced to the tip or ring of a different cable pair. Without the radar test set, the location of a split is often very difficult to find, especially on buried cable. The oscilloscope presentation for a split will show a capacitance discontinuity similar in form to that of a splice at the point where the conductors are separated. That is, a decrease in capacitance will be indicated by a positive pip as shown in Figure 6. If a re-split (restoration to normal condition) occurs in a subsequent splice, the pip will appear with the opposite polarity of that shown for the split.

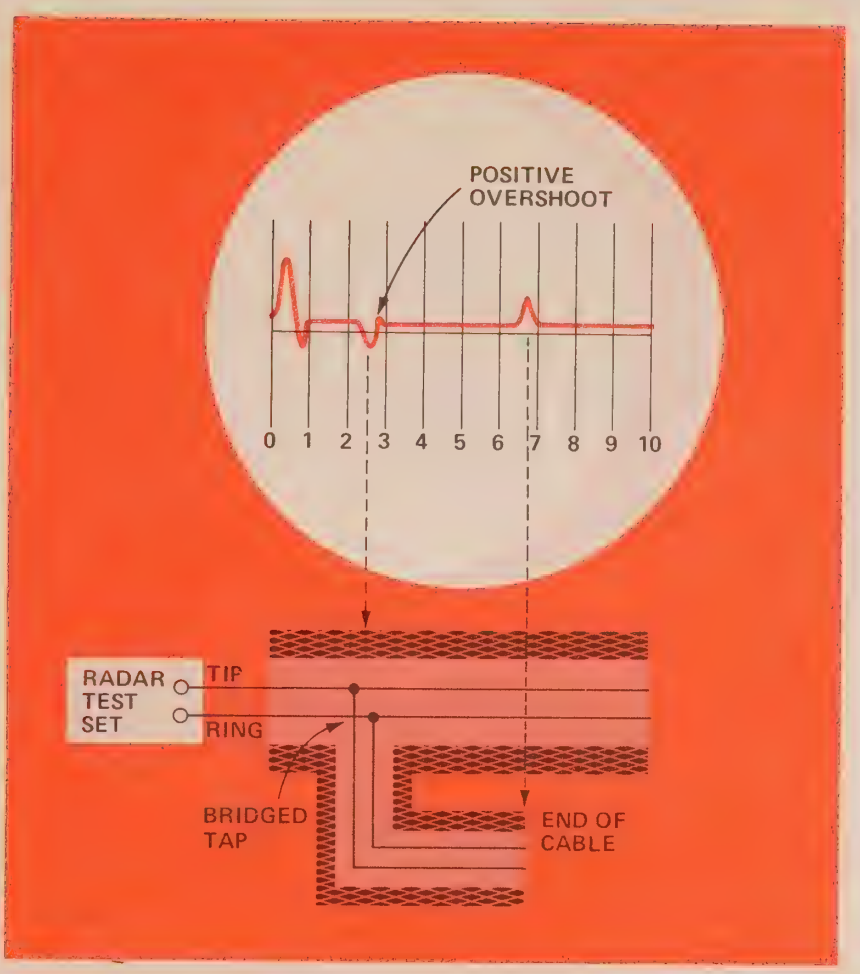


Figure 7. Bridged taps appear as negative pips with a positive overshoot.

Bridged Taps

For any type of FDM or PCM carrier operation, bridged taps must be removed since they represent a capacitance mismatch at high frequencies. These taps appear as negative pips, having a negative amplitude with a following positive overshoot or tail as shown in Figure 7. The amplitude of the overshoot is directly proportional to the length of the tap. Generally three or four taps with an average length of 100 feet will absorb all the output energy of the test set. In order to verify that all bridged taps are removed, it is necessary to repeat the test from the location of the last removed tap. A bridged cable and the main cable will present an overlapped trace. If a fault (short, cross or ground) also exists on the same cable pair, measurement from two locations is necessary to determine positively the location of the fault.

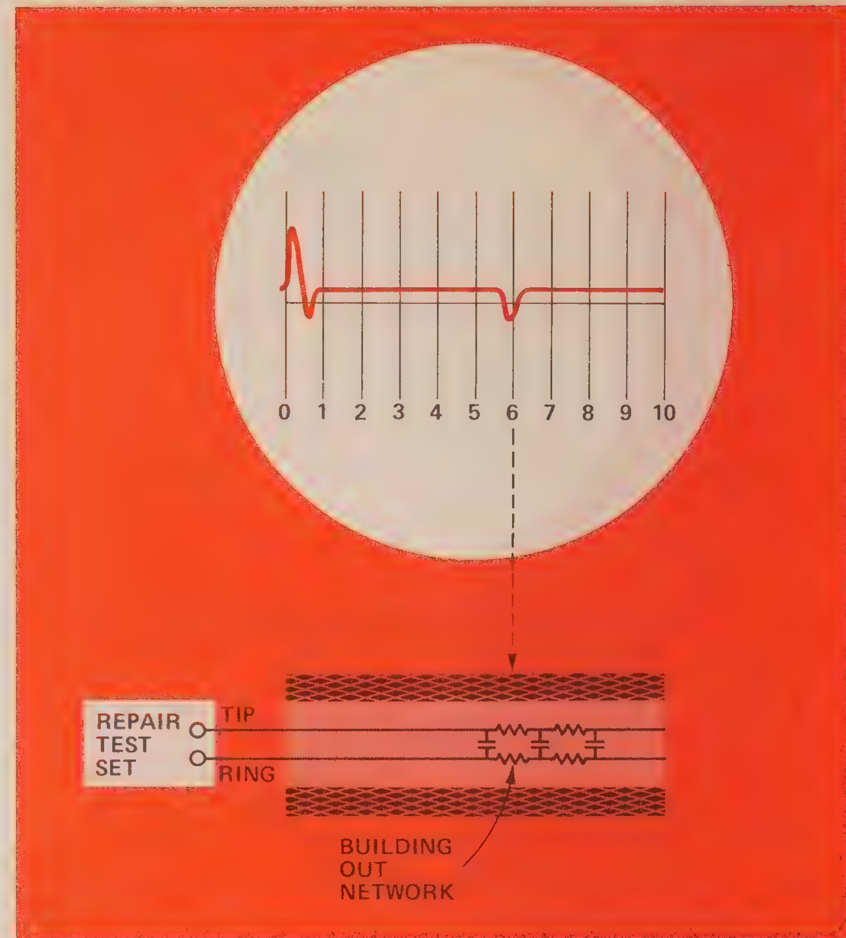
Loading Coil and Building-Out Network Detection

Loading coils and building-out networks that have been overlooked in the process of deloading a cable for carrier use, are perhaps the most difficult things to locate without the aid of an instrument such as a radar test set. A building-out network will appear on the oscilloscope display as a short circuit (the capacitors in the network short-circuit the radar pulse) and a loading coil appears as an open circuit since the pulse will not pass through the coil (see Figures 8 and 9).

Change in Wire Gauge

An increase in wire size such as at a splice decreases the resistance of the

Figure 8. Building-out networks appear as short circuits.



line and will give a negative pip similar to that of a bridged tap, but of lesser amplitude. The propagation velocity constant varies by approximately 1% per gauge number. (A change from 22 to 24 gauge will change the PVC by about 2%.)

Change in Dielectric Material

A change from PIC to PULP (paper insulation) cable will also cause a reflection indication. In this case, the propagation velocity constant will increase, producing a decrease in surge impedance, thereby yielding a negative pulse on the oscilloscope screen. The resultant pip amplitudes are generally small, less than those caused by bridged taps. If the dielectric constant of the insulation is known, the change in propagation velocity constant can be calculated by the formula:

$$\Delta PVC = \frac{1}{\sqrt{e}}$$

where e is the dielectric constant.

DC Loop Resistance and Conductor Resistance Balance

This test is advisable to verify the correctness of loop resistance and conductor resistance balance, which is necessary for proper carrier operation. Because the allowable resistance variation between conductors in the conductor resistance test is a maximum of ±0.5%, it is necessary to use a Wheatstone bridge since its accuracy is dependable within this tolerance. Measured loop resistance should check within 10% of the calculated values given in Figure 10. Due to the relatively wide tolerance in the loop test, and the fact that the balance test is a comparison measurement, it is permis-

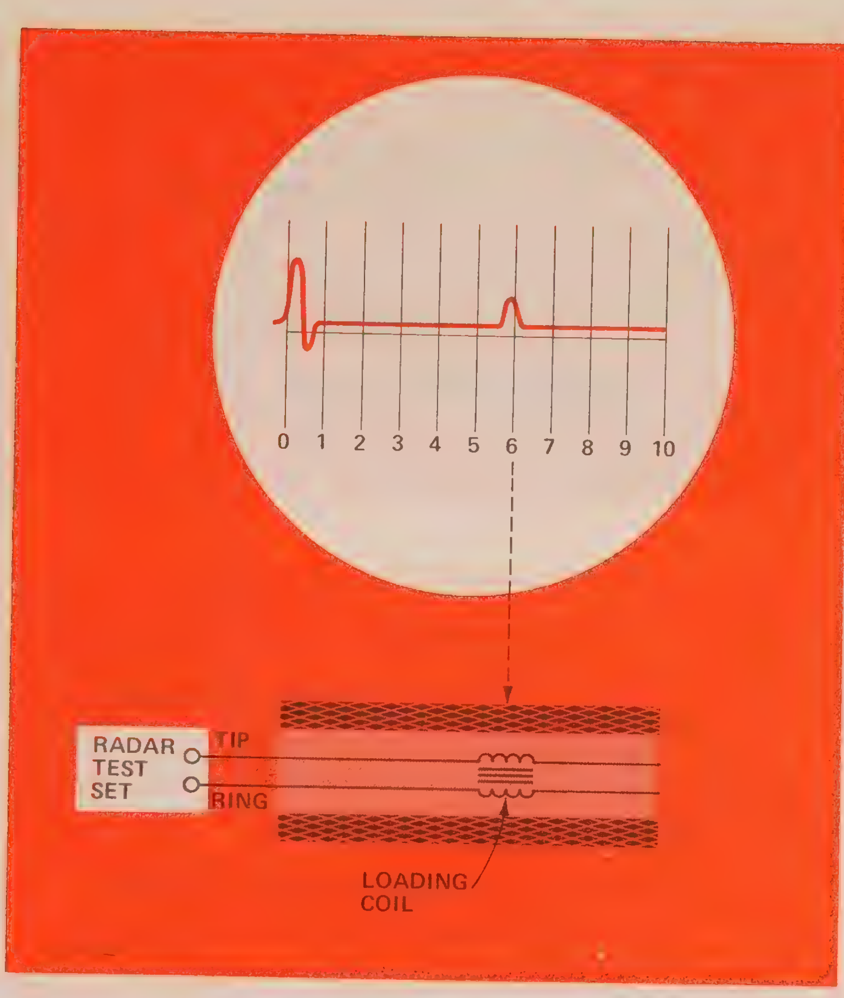


Figure 9. A loading coil appears as an open circuit.

sible to neglect the effect of temperature on the wire resistance.

With a Wheatstone bridge connected to the pair to be measured and a short-circuit placed across the distant end of the pair, the hookup will appear as in Figure 11A. The loop resistance should be within the required 10% tolerance of the calculated values given in Figure 10, noting that this is the total resistance "out and back." Removal of the short circuit at the distant end should indicate an open circuit. This step is necessary to verify that the pair being measured actually had a short circuit placed on it, and was not showing a loop due to some other connection.

Cable Gauge	Resistance Ω/kft at 68° F
16	8.03
19	16.10
22	32.28
24	51.34
26	81.62

Figure 10. Total "out and back" conductor loop resistance.

Example:

Assuming that 6200 feet (6.2 kft) of 22 gauge cable is used in one regenerator (repeater) section,

Loop resistance
= 6.2 kft x 32.28 ohms/kft
= 200.14 ohms.

Figure 11A. Loop resistance measurement.

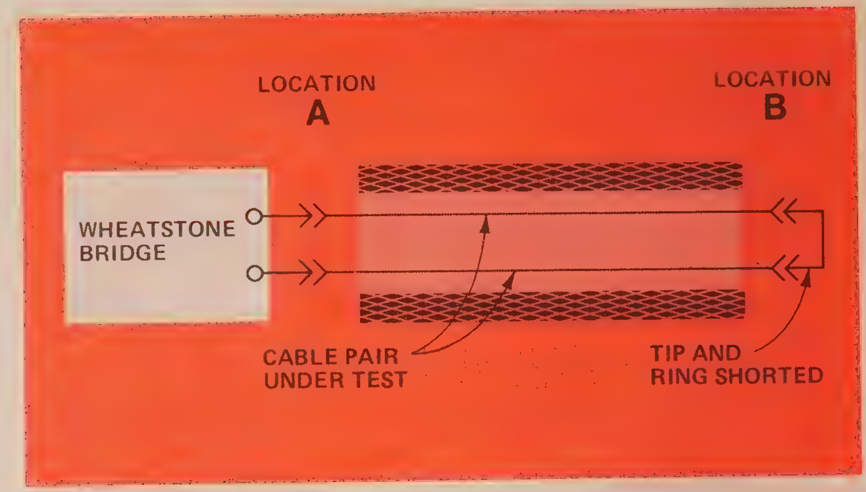
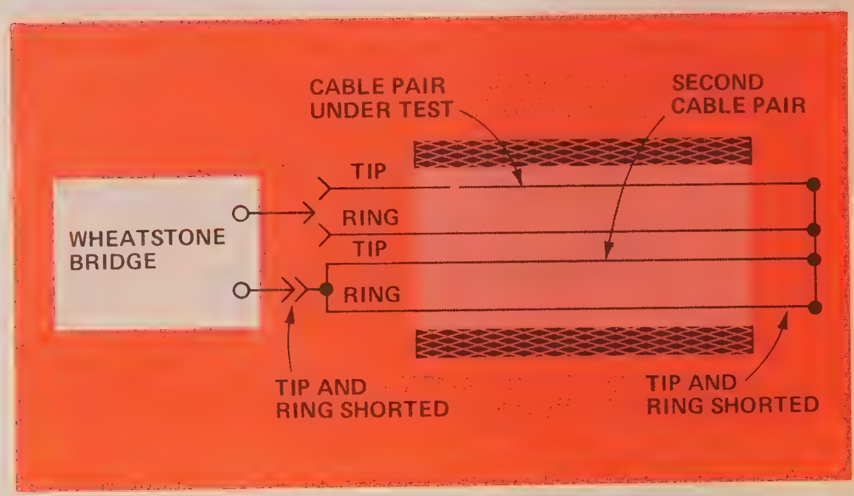


Figure 11B. Conductor resistance balance measurements.



To measure the conductor resistance balance, the Wheatstone bridge should be connected as shown in Figure 11B, using a second pair in the same cable as a third conductor. Individual measurement of the tip and ring conductors of the pair under test is necessary and is performed as shown by the arrow connections. First the tip then the ring of the pair under test is measured using the second cable pair as a return path.

Example:

For 6.2 kft of 22 gauge cable,

Loop resistance

= 6.2 kft x 32.28 ohms/kft
= 200.14 ohms

Average tip or ring resistance

$$= \frac{200.14}{2} = 100.07 \text{ ohms}$$

Average tip or ring resistance plus third-conductor resistance (1/4 of loop resistance)

= 100.07 ohms + 50.03 ohms
= 150.10 ohms

0.5% Tolerance in ohms

= 150.10 ohms x .005
= .75 ohms

Therefore, total resistance of tip or ring plus the resistance of the second cable pair should be between 149.35 ohms (150.10 - .75) and 150.85 ohms (150.10 + .75).

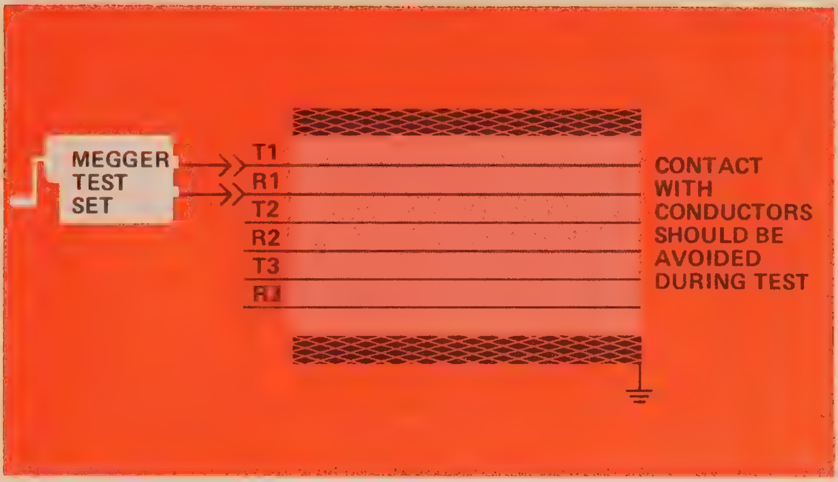


Figure 12A. Megger test on one cable pair.

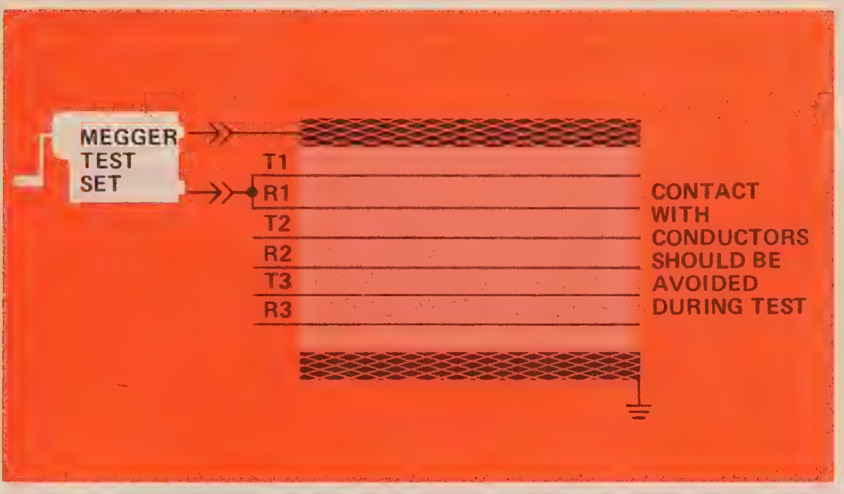


Figure 12B. Megger test of cable pair to ground.

Insulation Resistance Measurement

A megger test set is used to measure the insulation resistance between each conductor and ground and between each conductor of the pair. The distant end of the pair should be open and ungrounded for this test.

The operator should exercise extreme caution in avoiding contact with any of the terminals or wires during this test since the circuit is at a potential of several hundred volts above ground.

Once the megger is put into operation the resistance is then indicated on the megger test set meter. This resistance reading is then multiplied by the length in miles of the cable under test. The insulation resistance should be the same between the two conductors of a pair (Figure 12A) and between the pair to ground (Figure 12B). Paper insulated cable should indicate a minimum insulation resistance of 500 megohms per mile. Polyethylene insulated cable should indicate a minimum insulation resistance of 1000 megohms per mile.

Frequency Response

For PCM use, the cable pair frequency response must be checked to at least 1 MHz. Although the maximum energy of the PCM band occurs around 772 kHz, a considerable portion of energy exists up to 1 MHz. The frequencies above 1 MHz decrease in importance, so that measurements above this limit are not necessary.

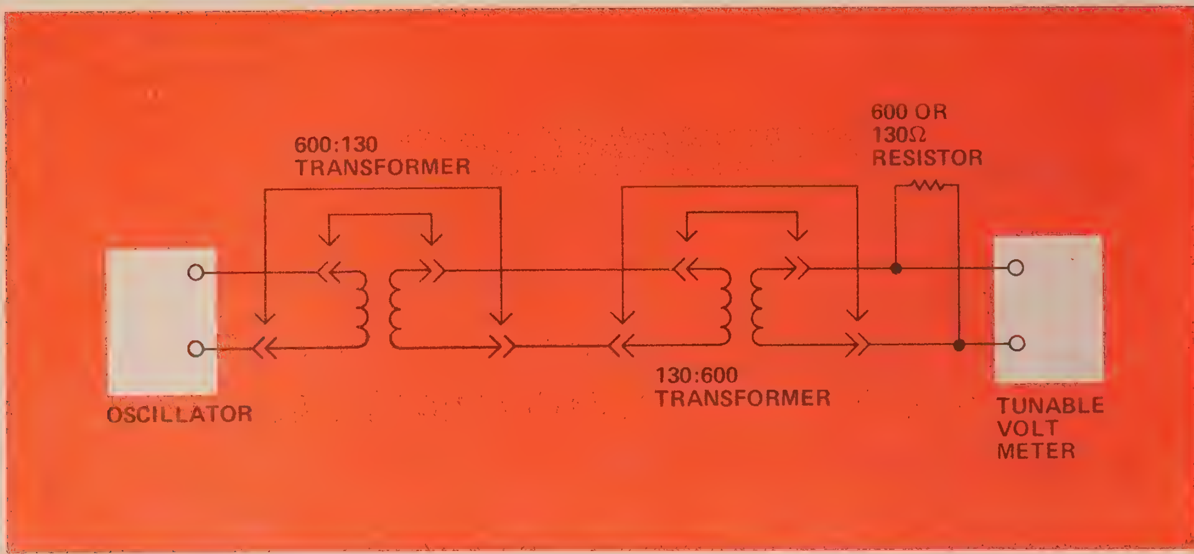


Figure 13A. Equipment arrangement for calibration.

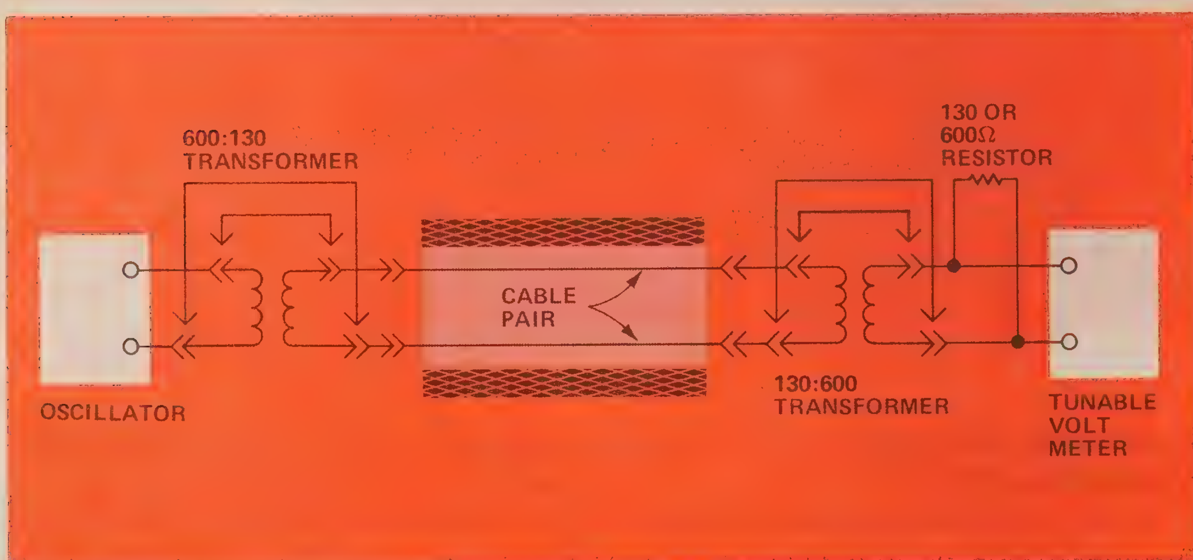


Figure 13B. Equipment arrangement for cable frequency-attenuation test.

Likewise, the frequencies below about 400 kHz decrease in importance for transmission of PCM circuits, so they can therefore be neglected during measurements.

To verify the frequency response and insertion loss of the test transformers, test equipment should be connected as shown in Figure 13A. The test transformers are necessary if the output impedance of the oscillator and/or the input impedance of the voltmeter is 600 ohms. At the frequencies of interest, the cable impedance is approximately 130 ohms. The oscillator frequency should be set to 1 MHz and the meter tuned to this frequency. The oscillator output level should be adjusted to indicate 0 dBm on the meter.

Tuning the oscillator between 400 kHz and 1 MHz, and tracking the frequency with the voltmeter tuner, any variation about the 0 dBm level, at

every 50-kHz interval should be noted and recorded.

At the end of the test the oscillator output power must still be 0 dBm on the meter. The oscillator output level during the frequency run on the cable must not be changed.

To check cable attenuation the test equipment should be connected as shown in Figure 13B. The oscillator should be tuned from 400 kHz to 1 MHz, and the received level reading should be recorded at every 50 kHz interval. In addition, the received level at 772 kHz should be recorded. This frequency is used in transmission calculations for the PCM carrier, and can be used to verify the cable attenuation figures used for calculation.

The amount of progressive attenuation of the cable should be plotted at the end of the test and any abrupt change in received level should be noted since this would indicate a change in the transmission characteristics of the cable. Paper-insulated cable will usually exhibit a somewhat higher attenuation than PIC, perhaps 0.5 to 1.0 dB greater per thousand feet at 772 kHz.

Economic Benefits

Returning to the original question – what is at fault on a system with new equipment and older cable, when only part of the total number of channels function properly? With pulse reflection, resistance, and frequency response tests complete, and with any unserviceable pairs removed, this question may now be answered.

Once equipment checkout is complete and verification of adequate crosstalk coupling loss in the system is made, the communications equipment user may then reap the economic benefits of using previously installed cable for his new equipment.

GTE LENKURT

DEMODULATOR

DECEMBER 1972

Phase Jitter

Until recently, phase jitter was little more than a definition to the communications industry. Since phase jitter did not cause errors in voice or low-speed data communications, there was little need for a comprehensive understanding of the subject. But, with high-speed digital transmission, phase jitter as slight as 1° may cause errors in the reception of the bit stream.

Transmission engineers have been aware of the existence of phase jitter for some time, but since it has little effect on voice intelligibility, it was not considered a significant measure of telephone channel performance. Today with the expanding use of the telephone network for the transmission of high-speed digital signals and the increased use of pulse code modulation (PCM), phase jitter, or time jitter, has become a channel performance parameter of increasing concern. The transmission channel has not changed but phase jitter is more deleterious to high-speed signals than to voice signals with their redundant nature (only about 25% of a voice signal is needed for intelligibility).

As transmission speeds increase, data pulses become narrower and closer together. Eventually, these pulses become so close together that phase jitter can displace a pulse enough that the receiving equipment "sees" a pulse where it should see a space, or vice versa. Digital transmission is sensitive to phase or time jitter regardless of the modulation technique used – phase, frequency, or amplitude. Consequently, jitter is now a matter of increasing importance to the equipment designer, transmission systems engineer, and data systems user.

In recognition of phase jitter's increased significance, transmission performance objectives and measurement standards have been set by a number of groups such as D.C.A. (Defense Communications Association), Bell System, and E.I.A. (Electronic Industries Association). It is particularly important that the definition and measurement of phase jitter be thoroughly understood by the designer, engineer, user, and technician in the light of its increased importance in the telecommunications field.

In high-speed digital transmission, phase jitter is used to describe the change in periodicity that results during transmission. And, time jitter refers to synchronization errors on PCM transmission systems.

Phase jitter on data transmission systems has been traditionally defined as unwanted change in phase or frequency of the transmitted signal due to modulation by another signal during transmission. Using this definition, if a single-frequency test-tone is frequency or phase modulated during transmission, the received tone will have sidebands associated with it (a single-frequency is used for testing and measuring phase jitter). The amplitude of these sidebands, compared to the received tone, is a measure of the phase jitter imparted to the signal during transmission (see Figure 1). This phase jitter can be the result of intersymbol interference (distortion caused by the tails of preceding and

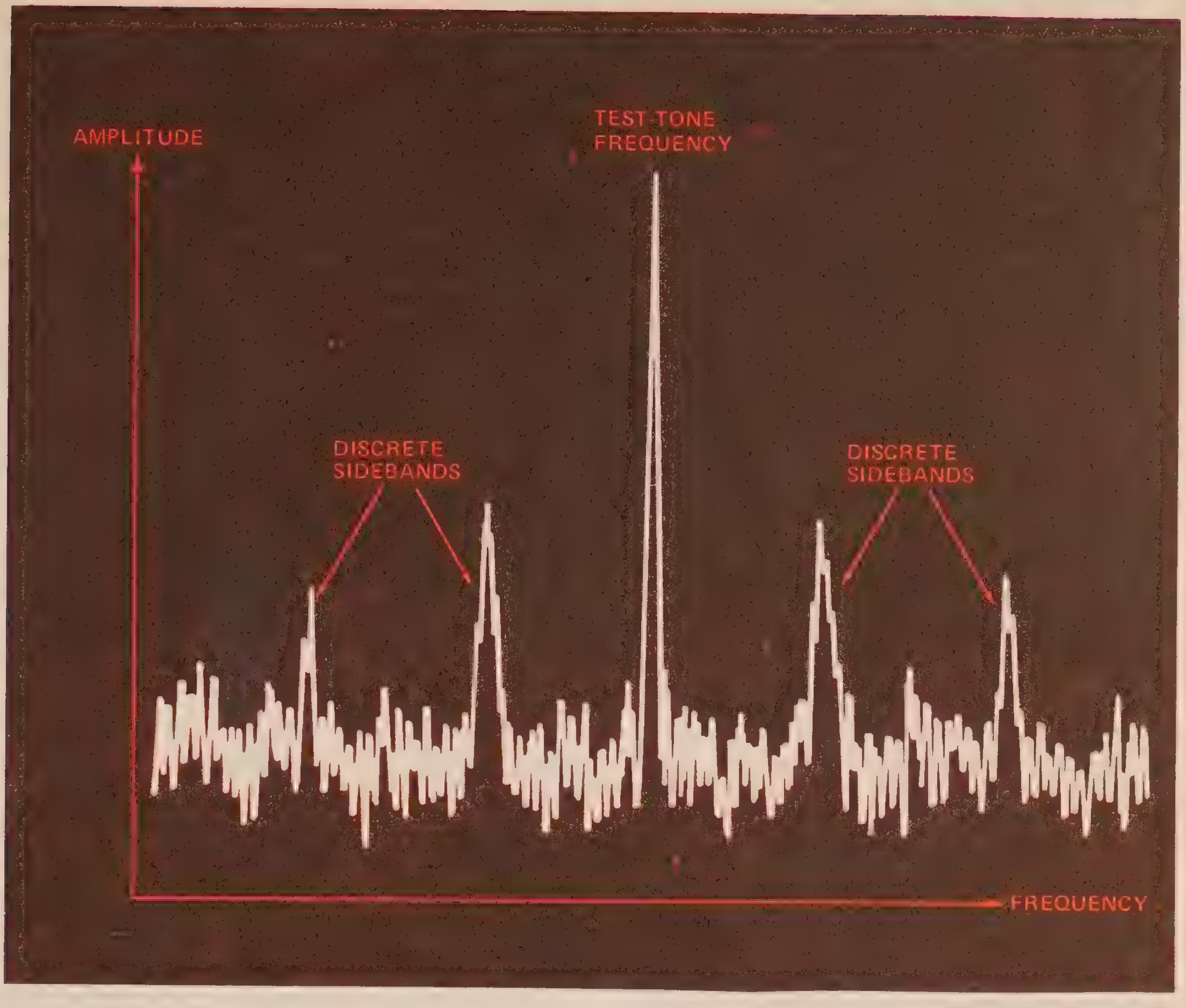


Figure 1. When testing for phase jitter, a single-frequency test-tone may appear to have a definite spectral distribution such as symmetric, discrete sidebands or it can have random, low-amplitude sidebands like noise.

succeeding pulses into the time slot of the pulse currently being transmitted) or crosstalk (interference caused by energy being coupled from one circuit to another). Phase jitter is measured in degrees of peak-to-peak variation for each cycle of the transmitted signal. And, time jitter is measured in seconds of variation for each sampling period.

The definition of phase jitter as a modulation process specifies its characteristics and distinguishes it from the amplitude/phase variations caused by other forms of distortion such as spurious signals, crosstalk, impulse noise, and channel noise.

Phase jitter can also be defined as *any* unwanted variations in the zero crossings (when the transmitted signal returns to zero amplitude) of the received signal. Data modems look at zero crossings so this is an easy characteristic to measure. Figure 2 illustrates how the zero crossings can vary on a digital signal that has phase jitter.

In a carrier channel, if a single-frequency test-tone is modulated by a second pure tone, sidebands are generated at the carrier frequency plus and minus the modulating frequency of this second tone. The effect of these sidebands depends on the modulation process. If the sidebands amplitude modulate the transmitted signal, the signal amplitude varies, but its phase is not affected by the sidebands. There-

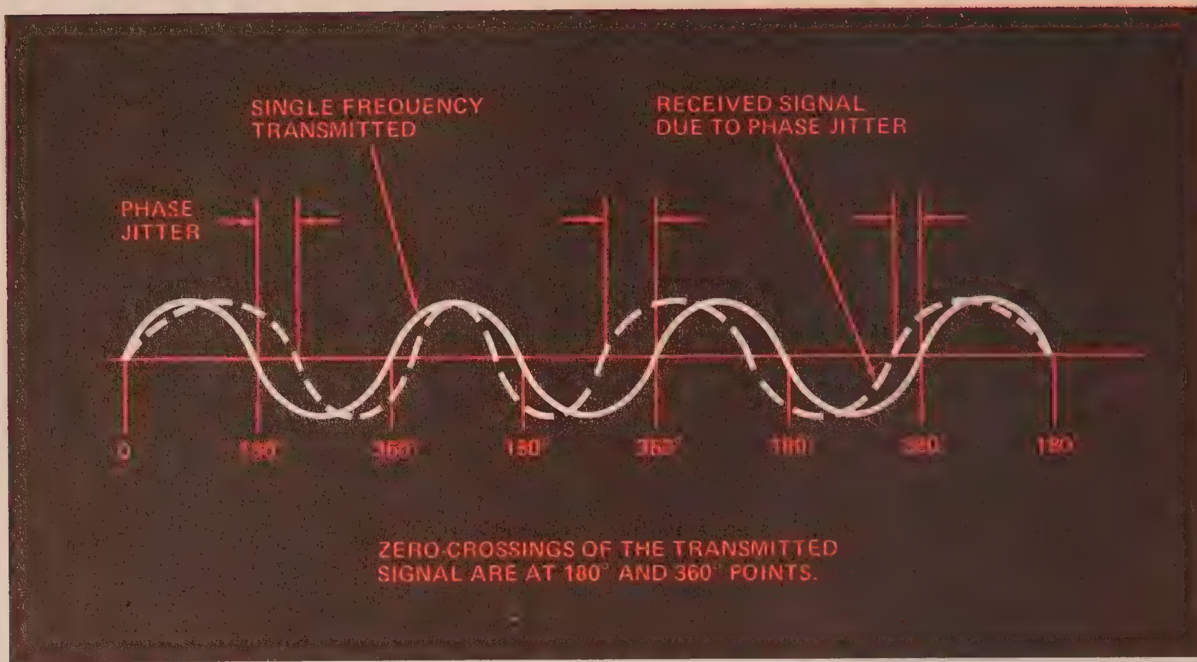


Figure 2. Phase jitter varies the zero crossings of the received signal.

fore, the periodicity of the carrier's zero crossings is unaffected and no phase jitter is produced. During phase modulation, the amplitude of the modulated wave remains constant, while the phase varies with the amplitude of the modulating signal. With phase modulation, the sidebands are always 90° out of phase with the transmitted signal and will therefore phase modulate the signal, changing the periodicity of the carrier's zero crossings — causing phase jitter.

A single interfering tone can produce both angle modulation and amplitude modulation. Each additional interfering tone acts not only on the carrier, but on each of the other tones as well. Many such interfering tones spaced at random frequencies approximate white noise. Therefore, noise and other types of interference are inextricably tied in with any jitter measurement.

Sources

The modulation process that causes phase jitter may be either phase or frequency modulation depending on the source which is generally in the terminal equipment and not generated on the line. It generates symmetrical sideband components spaced at the jitter frequencies (the transmitted frequency plus and minus the frequency of the modulating signal), and the amount of jitter is a function of the magnitude of the jitter source and not a function of the frequency or level of the transmitted signal.

Undesired incidental phase modulation of digital signals is of the most concern for those signals transmitted over carrier circuits. In multiplex units, phase jitter results from incidental phase modulation of the oscillators used for the translation of the signals within the channel to a different part of the spectrum for transmission. This incidental phase modulation is caused by noise, and line-related ripple on office batteries and on power and bias supplies, and within timing circuits on FDM systems. This modulation is transferred to the multiplexed signal during frequency translations and generally, the greater the number of translations on a circuit, the greater the

phase jitter. Therefore, the signals transmitted over channels in the higher supergroups have the most phase jitter.

Measurements have indicated that for long-haul, multiple-link systems, the most serious components of incidental phase modulation are power-line related, both as harmonics of the line frequency as well as fractions of the line frequency. With an individual carrier system, the predominant components of the phase modulated signal are low-frequency, time-varying signals, which are quite often related to noise occuring on office power supplies.

Thus, phase jitter most commonly occurs at rates related to the power-line frequency (60 Hz) and its harmonics and submultiples, the telephone ringing frequency (20 Hz), and the interaction between these two sources. Modulation components above 300 Hz rarely occur, if they do, they are accompanied by large amounts of jitter below 300 Hz.

Although phase jitter principally occurs in carrier equipment, it may also be caused by automatic equalizers and associated equipment.

Additive Signals

The effect of channel noise, typical of the undesired parameters that the channel adds to the test tone, is distributed throughout the circuit bandwidth. These additive signals cause simultaneous variations in both amplitude and phase. This occurs whenever two or more signals are added. These amplitude/phase variations are completely related and either may be measured to completely describe the parameter. The amplitude variations are more easily measured, and these additive signals are typically described in terms of their levels — noise level, distortion level, or spurious tone level.

The only parameter which varies just the signal phase is angle modulation. It is this incidental angle modulation (which includes both phase and frequency modulation) that is defined as phase jitter. The magnitude of the signal-correlated noise is a function of the signal level, the signal-to-noise ratio therefore remains essentially constant when the test tone level is varied. For this reason, jitter readings typically show little change for a large variation in test-tone level.

Extensive measurements have been made to determine if phase jitter varies with test-tone frequency. Theoretically, a variation would not be expected as the phase jitter originates on the carrier supply oscillators and is then uniformly transferred onto the signals in the channel during frequency translations. Variations which do exist have been shown to be caused by the modulation of the jitter sideband structure, by the channel amplitude and envelope delay characteristics and by the additive effects of crosstalk, spurious tones, and unequal noise distributions which result in a variation of the additive signal effects as the test-tone frequency is varied.

Measurements made at the standard test-tone frequency represent channel behavior and provide standardization of measurement, utilize existing test-tone generators and eliminate variations in readings due to channel effects not related to phase jitter.

Phase Jitter Measurement

Phase jitter measurements are made by transmitting a single-frequency tone on the circuit of interest and observing the frequency distribution of the received signal. The standard test-tone frequency is 1020 Hz (1000 Hz is no longer used since it is a submultiple of the 8000-Hz sampling frequency of PCM systems).

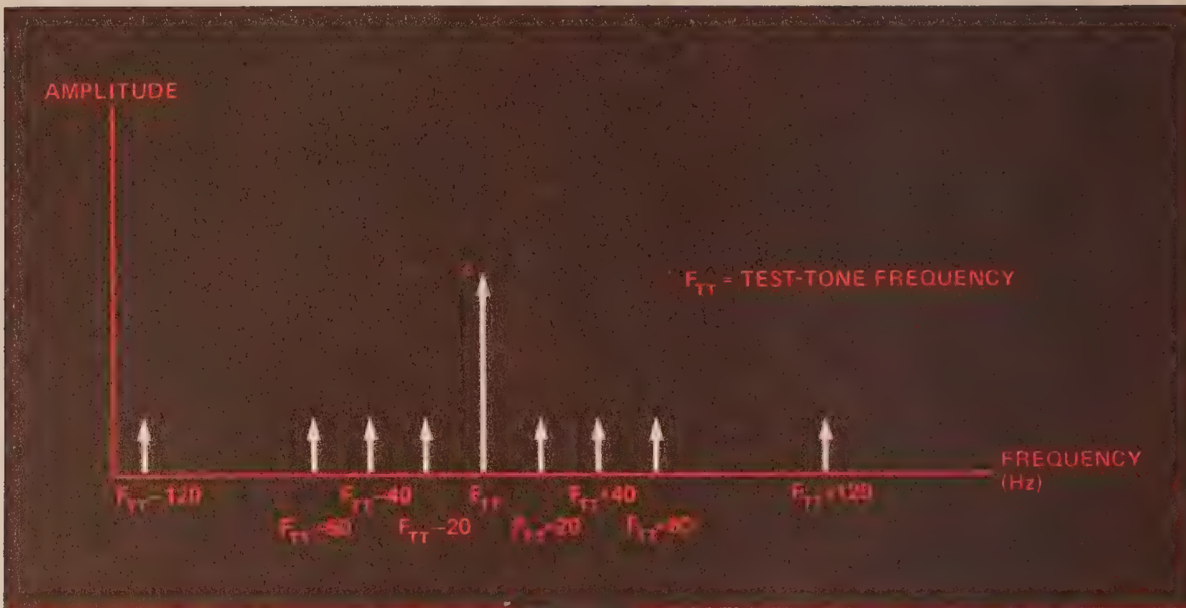


Figure 3. Phase jitter caused by power-line frequencies and telephone ringing frequencies have a definite spectral distribution.

The frequency distribution of phase jitter is shown in Figure 3. The sideband structure shown is similar to that of an amplitude modulated carrier, and differs only in the phase relationships between the carrier and its sidebands. Thus, the measurement technique used must be able to differentiate between amplitude and angle modulation.

There are several methods of measuring phase jitter. The oscilloscope method provides a real and meaningful measurement in the lab, however, it is not necessarily suitable for field measurements. The oscilloscope must have performance capabilities which are commensurate with a good lab oscilloscope and may not always be found in the type of scope used in the field. In addition, relatively unskilled personnel would most likely find the use of the scope and resultant interpretation of the peak-to-peak readings difficult.

Another technique makes use of the measurement of the carrier-to-sideband level, then relates the measurement to peak-to-peak phase jitter via a chart. This is perhaps a more convenient and conventional measurement that is made throughout the telephone industry. Although the equipment used is complex and expensive, it is more familiar in the telecommunications field.

Still another technique involves the use of two test tones and has the advantage of considerable operational simplicity for the technician making the measurement. This technique has the possible disadvantage that a special generator must be provided, while the other techniques use standard readily available test signals.

A more convenient measurement technique uses zero-crossing detectors to check for disturbances in the periodicity of the received signal. There are many factors that cause unwanted zero crossing variations. The measurement technique used must give some indication as to the jitter source causing this variation. Unless this source information is given, there is little to be gained by making the measurement.

The frequency distribution of the jitter components is an important characteristic which permits at least

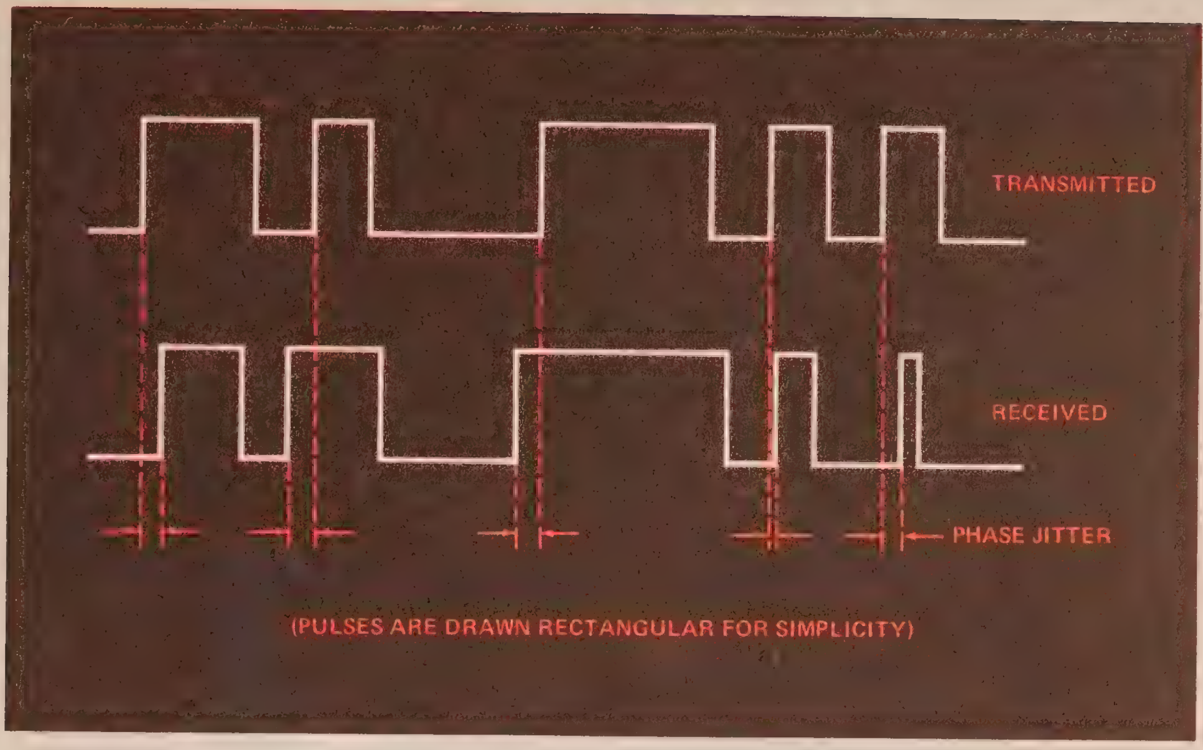


Figure 4. Phase jitter causes the received digital pulses to change in duration and can cause errors when translating pulses and spaces at the receiving end.

one type of jitter to be distinguished from another. The traditional hum modulation caused by power-supply ripple is relatively easy to identify since it nearly always produces jitter components that fall within 20-300 Hz of the single-frequency tone used for testing and measuring phase jitter. Other jitter sources produce components that are generally distributed throughout the circuit bandwidth.

PCM Systems

Hum modulation is not the source of jitter on PCM systems, but noise, near-end crosstalk (NEXT), and intersymbol interference are the major causes of PCM time jitter. Signal distortion in PCM transmission results from many factors – sampling, quantizing, and compandor mistracking – all of which can be referred to as constant power distortion.

In contrast to the measurement of phase jitter, a small change in the test-tone frequency can make a substantial change in the distortion spectrum when measuring jitter on a PCM system. Time jitter is frequency dependent. As the test-tone frequency is changed, the jitter changes in both magnitude and spectral distribution. However, time jitter is independent of the level of the test tone.

Time jitter can be minimized through careful equipment design and system engineering. A given time displacement has a greater effect on the shorter period of high-frequency signals (see Figure 4). Thus, time jitter becomes more critical with increasingly higher bit speeds.

Minimization

In designing digital transmission equipment, care should be taken to select coding methods, modulation techniques and other parameters that will minimize the effects of phase jitter.

GTE LENKURT

DEMODULATOR

JANUARY/FEBRUARY 1975

PCM Update

PARTS 1 AND 2

NEW *DEMODULATOR* CIRCULATION PROCEDURE

Since the *Demodulator* was established in 1952, it has shown a steady increase in subscribers. By the end of 1974, the number of subscriptions, including both foreign and domestic, had reached 45,450 (the Spanish translation has an additional 5,000 subscribers). It is unfortunate that, while the number of subscribers has steadily increased, so also have the costs of producing and circulating the *Demodulator*, especially the mailing costs. Because these are times when economizing is a must in practically any endeavor, a decision has been made to modify the *Demodulator* circulation procedure.

Beginning with the January and February issues of the *Demodulator*, two monthly articles will be combined in one mailing so that there will be six, rather than twelve, mailings a year. In this way, mailing costs will be substantially reduced, but the number of monthly articles published yearly will remain the same as before. GTE Lenkurt hopes that this economizing step will not be a great inconvenience to the readers of the *Demodulator*.

Jose C. de Leon, Editor
The GTE Lenkurt Demodulator

PCM Update - Part 1

Pulse code modulation techniques have established a significant foothold in the field of communications. Here is an update on some of the recent developments, concepts, and possibilities in this field.

Pulse Code Modulation (PCM) was conceived by Alec H. Reeves in 1937 and patented in 1938. Although some of its potential as a telecommunications technique was recognized then, the concept nevertheless remained dormant a whole decade for lack of the necessary electronics. With the development of high-speed, solid-state switching devices shortly after World War II, the possibility of a working PCM system stirred interest in the telecommunications industry, and 1948 saw an increase in PCM studies. Still, it wasn't until the mid-fifties – when the use of semiconductors became economically feasible – that PCM became a practical possibility. The first commercial U.S. system was installed in Akron, Ohio, in 1962 for Bell Telephone of Ohio.

The original Western Electric T1 carrier system, which included the D1 PCM channel bank, was conceived as a short-haul, heavy-route carrier system designed to relieve cable congestion in metropolitan areas. The T1 system offered the reliability and quality of voice transmission necessary for use on direct trunks between Class 5 end offices and toll-connecting trunks, or tandem trunks between Class 5 end offices and Class 4 toll or tandem offices (see Figure 1). After the introduction of the T1 system by the Western Electric Company, many independent manufacturers began to produce similar T1-type equipment.

In spite of its popularity, however, the D1-type channel bank did not have the necessary transmission quality to be widely applied in the intertoll networks, where there could be as many as seven trunks in tandem, from an originating Class 4 office to offices 3-2-1 and back down 1-2-3 to the terminating Class 4 office. The D2 channel bank was subsequently designed and introduced into the telephone networks, particularly for service in the intertoll network (see Figure 2). Besides better transmission characteristics, it offered features tailored to intertoll and special service usage, and also offered some maintenance and alignment simplifications. Independent manufacturers again produced equipment that was end-to-end compatible with the D2 channel banks.

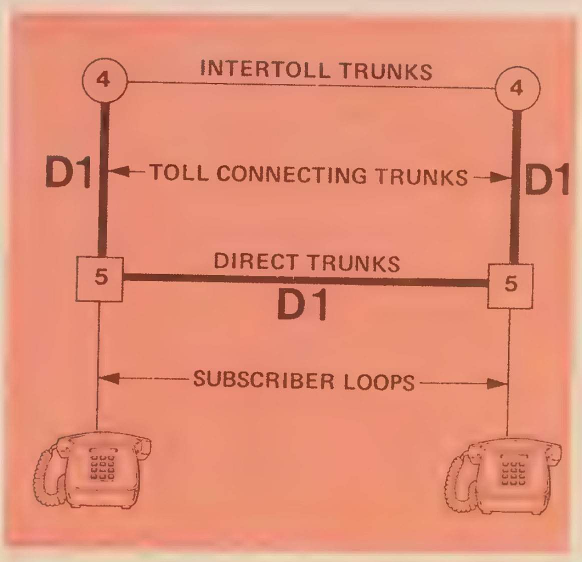


Figure 1. D1 channel bank service.

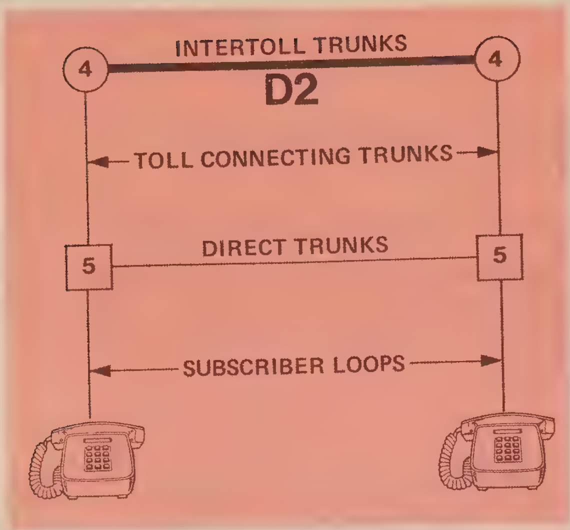


Figure 2. D2 channel bank service.

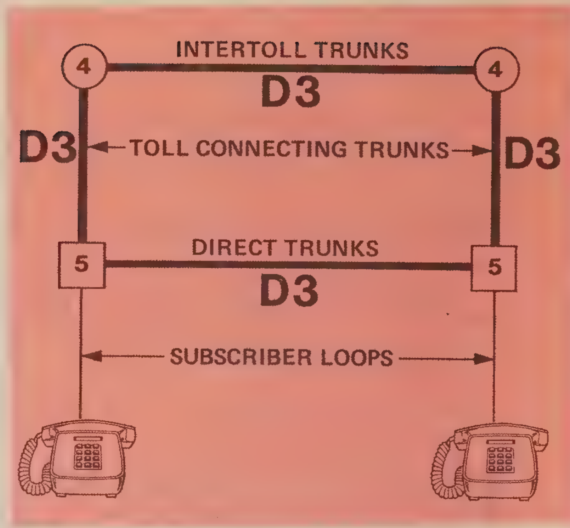


Figure 3. D3 channel bank service.

While the D2's higher quality transmission performance would allow it to be applied to direct or toll connecting trunks, its intertoll service features made it too expensive to apply in the lower steps of the network, a condition which led to the further development of the D3 channel bank. The D3 channel bank was originally conceived as a technological update of the D1 channel bank, to be used in the same type of service. However, it was soon realized that, with the new technologies available, such as integrated circuits, there was little or no cost penalty in making the D3 operation compatible with D2. This would afford greater flexibility for D3, allowing it to be used in intertoll service as well as direct and toll connecting service (see Figure 3). While channel bank designations D1, D2, and D3 are Western Electric designations, they have almost become generic terms used to describe D-type compatible equipment manufactured by independent telephone companies.

The Western Electric D2 is packaged as a 96-channel terminal with some circuits shared over all 96 channels. But electrically it functions as four nearly independent 24-channel terminals. Each 24-channel transmission package, called a "Digroup," operates over a separate 1.544 Mbs T1 line. The independent manufacturers, however, have all designed D2-compatible channel banks as 24-channel terminals, completely self-contained. These terminals all have the higher transmission qualities of the D2, and features applicable to intertoll and special service usage.

The W.E. D3 was also designed as a 24-channel terminal, completely self-contained, with the same transmission qualities as the D2. Besides the 96- versus 24-channel package, the main differences between D3 and D2 are different channel sampling sequence, slightly different alarm sequence, and different maintenance features. Since the independent manufacturers' channel banks were already in 24-channel packages, very little was required to make them completely end-to-end compatible with either W.E. D2 or D3. Most manufacturers have options available which give their channel banks D2/D3 capability. This makes these channel banks applicable in all types of trunk service because of the flexible maintenance and interface features in the various types of channel units.

PCM Subscriber Carrier

PCM subscriber carrier has been a development of independent carrier

manufacturers, with no equivalent offering, as yet, from Western Electric. Consequently, no industry-wide standardization has been set, and some manufacturers offer PCM subscriber carrier based on their D1, while others offer subscriber carrier based on their D2/D3.

In this context, "subscriber carrier" is defined as a point-to-point system, in contrast to "station carrier," which is defined as a distributed system. Subscriber carrier allows the extension of central office service to a single remote point for distribution (see Figure 4). Station carrier allows the extension of central office service to several distributed points along the extension route. PCM "subscriber" carrier generally allows subscriber loops at its outward (or subscriber) end to be equal to or longer than the loops allowed at Class 5 end offices (up to 1600 ohms); other types of "station" carrier generally limit the subscriber loop to 200 ohms or less. Therefore, PCM subscriber carrier actually adds to the telephone network, for a possible total of 11 links instead of the traditional 9 links previously allowed between subscriber loops.

There is a possible disadvantage in using PCM subscriber carrier based on the D1 format as opposed to that based on the D2 format, since the D2/D3-based subscriber carrier has better transmission performance, thereby maintaining subscriber-to-subscriber performance at a higher level for longer distances. Another major advantage of the D2/D3-based PCM subscriber carrier is that its line format is compatible with digital time division multiplex (TDM) switching, which is to be introduced into Class 5 offices in the not-too-distant future.

TDM Switching

Time division multiplex switching equipment for Class 4 toll offices or tandem offices is now being developed by both Bell Labs (No. 4 ESS) and Automatic Electric (No. 3 EAX); these will be introduced into the telephone network sometime between 1977 and 1980 (see Figure 5). These switching centers operate at 1.544 Mbs to provide high-capacity switching from one PCM line to another. This is accomplished without demultiplexing from 1.5 Mbs down to voice frequencies for switching, and back up to 1.5 Mbs for

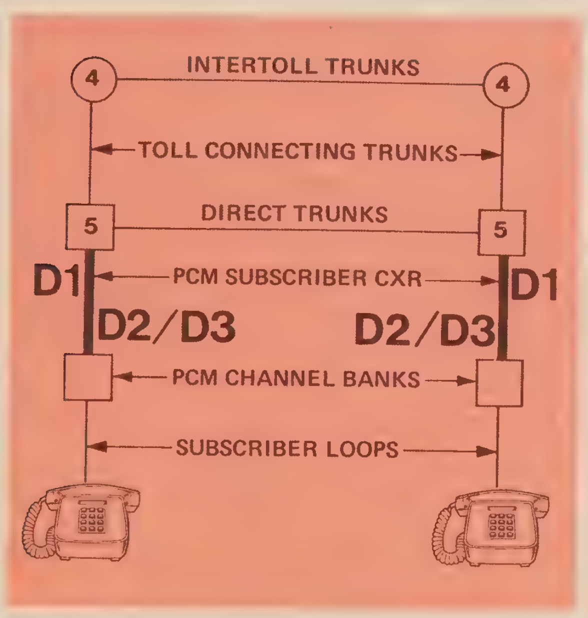


Figure 4. PCM subscriber carrier service.

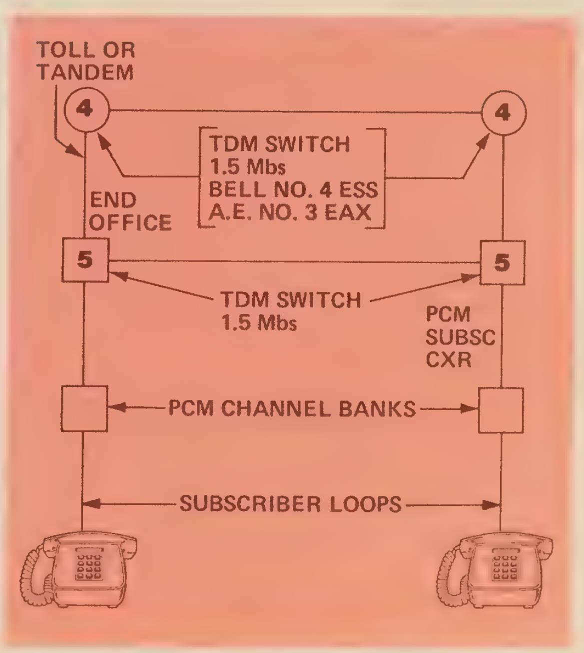


Figure 5. Digital switching capability.

re-transmission. Only the D2/D3 PCM 8-bit word format will be allowed in these switching centers, so connection of D1 channel banks directly to a switch at 1.5 Mbs is prohibited.

Soon after introduction of digital switches into the toll or tandem offices, TDM switches will be introduced into Class 5 end offices. It appears now that these switches will also allow only the D2/D3 format. Therefore, PCM subscriber carrier based on the D2/D3 channel bank has a distinct advantage.

Digital switching is accomplished by the switch breaking down the received 1.5 Mbs pulse stream into individual channel 8-bit PCM words, storing them until the appropriate channel time comes along in another addressed 1.5-Mbs pulse stream, and then transmitting them in the appropriate time slot (see Figure 6). Thus, information in a channel 2 slot coming from an end office to a No. 3 EAX may be switched to a channel 15 slot going to another office; this connection is maintained through the switch as long as the conversation lasts. However, the actual "connection" is only made at the appropriate time slot, leaving the 23 other time slots available for other channel connections.

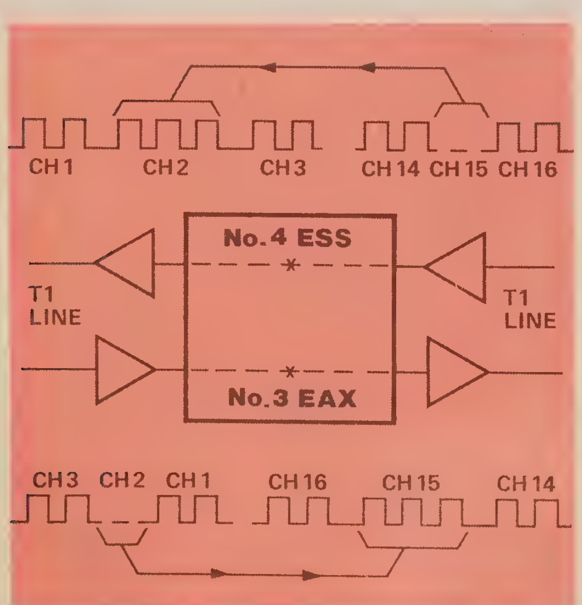


Figure 6. Toll or tandem TDM switching (8 bit PCM).

In addition to the voice transmission considerations for each of the types of telephone trunks, the signaling requirements of each must be considered. In the intertoll network, E&M signaling is most prevalent, with loop signaling capability also required. When common channel interoffice signaling (CCIS) is introduced into the network, a flexible PCM carrier must have the ability to accommodate this type of signaling. In direct or toll connecting trunk use, the most prevalent type of signaling is loop signaling, with some requirements for E&M and Foreign Exchange (FX).

A truly flexible PCM channel bank to be used for PCM subscriber carrier should be able to accommodate many types of signaling, including bridged, divided and superimposed ringing, paystation, and foreign exchange. Channel banks of the D2/D3 type provide improvements over the D1 channel banks, including a greater signaling flexibility, improved noise and distortion performance, and a standardized digital switching format.

PCM Format

The three basic operations necessary for PCM voice operation are sampling, quantizing, and encoding. Basic sampling theory indicates that an analog wave, such as a voice representation, can be reproduced with very little distortion if the sampling frequency is at least twice as high as the highest frequency to be transmitted. Since voice transmission of about 4-kHz bandwidth is to be transmitted in the telephone network, the sampling rate of 8 kHz has been standardized in the United States and is used for D1, D2, and D3 operation.

The number of steps used to represent the sample (or quantize it) determines the quality of the recovered

voice signal at the receive output of the system. Or, put another way, the number of steps determines the "quantizing distortion." While the D1's use of 128 quantizing steps was adequate for direct or toll-connecting trunks, the D2 (and D3) required 256 quantizing steps to provide adequate quantizing distortion for the intertoll network.

The number of binary bits necessary to give a digital code representation of 128 steps is 7, while to represent 256 steps requires 8 bits. Therefore, the D1 encoded "word" is a 7-bit word and the D2 or D3 word is an 8-bit word (see Figure 7).

The most important and most basic characteristic of a PCM system is its frame format. Once this is understood, the whole system can be understood. Since the sampling rate is 8 kHz, the time between successive samples of a given channel is 125 μsec. The "frame" length, then, is 125 μsec, during which all channels in the system must be sampled, quantized, and encoded in sequence. Then, the first channel is sampled again, and the sequence repeated (see Figure 8).

For the D2/D3 system, 24 channels with 8-bit words are included in one frame, giving 192 bits. Then a 193rd bit is added for framing and signaling control. Since it is "shared" between

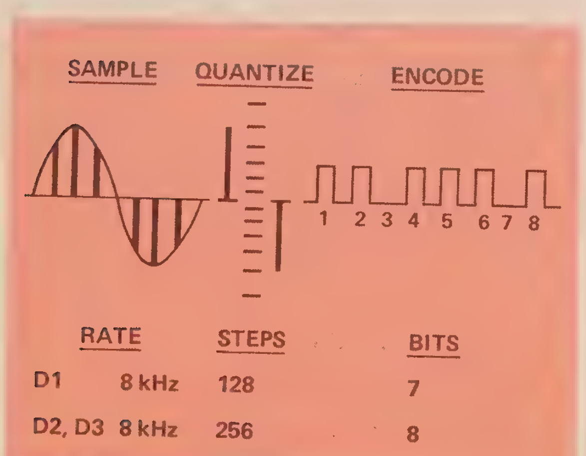


Figure 7. Pulse code modulation coding.

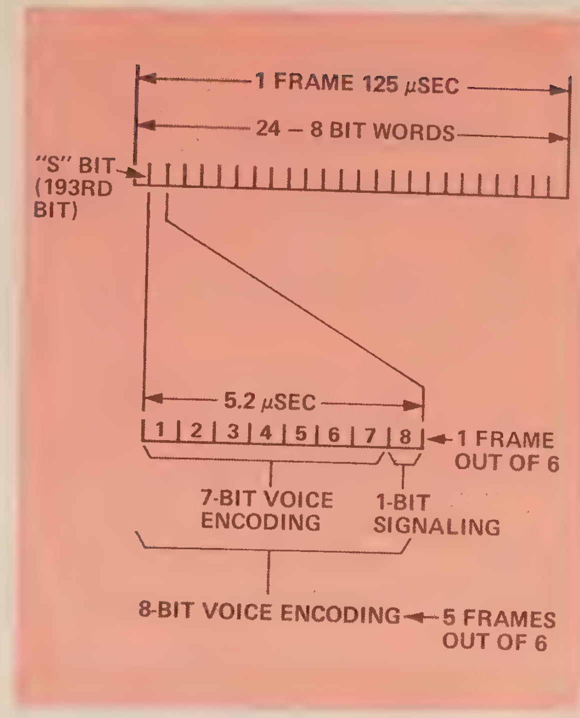


Figure 8. D2/D3 PCM frame format.

these two functions, it is called the "S" bit.

Each of the 8-bit words in the 125 μsec frame length is allowed 5.2 μsec. For the least amount of quantizing distortion, all 8-bits should be used for voice encoding. However, signaling for each channel (such as dial pulsing and supervision) must also be transmitted. But, since dial pulsing and supervision are of such low frequency compared to voice transmission, it can be sampled and transmitted at a slower rate than the voice. Therefore, a compromise is reached to transmit 8-bit voice samples for 5 frames out of 6, and in the 6th frame, transmit a 7-bit voice sample and a 1-bit signaling sample.

Future plans call for common channel interoffice signaling (CCIS), whereby signaling for all channels is carried by a single data channel. This will allow 8-bit encoding at all times, and improve quantizing distortion by about 2 dB.

Coding

The D2/D3 coding utilizes a nonuniform coder-decoder (Codec) trans-

fer characteristic (see Figure 9). That is, low level samples at the input of the coder are compared against (quantized by) small steps, whereas high level samples are quantized by large steps. This is a compression technique which maintains relatively constant quantizing distortion over a wide range of talker volumes.

For a given size of coder input sample, one size of decoder output sample is produced. However, for several samples close to the size of the first sample, the same decoder output sample will still be produced. This is quantizing distortion, in that not all samples can be reproduced exactly. But, by using small quantizing steps for small samples and large quantizing steps for large samples, the same relative error (in dB or %) can be maintained for small and large samples. Since this is a compression technique, a typical compression curve is described, where small input signals are reproduced nearly linearly, but large signals are attenuated to form the output compressed signal. Since this is a digital system, it is desirable to have a segmented compression characteristic rather than a continuous one. Therefore, from subjective talker-listener considerations, the curve is made up of 8 positive segments for positive samples and 8 negative segments for negative samples. The 8th segments, positive and negative, are co-linear, so the curve is called a 15-segment curve (see Figure 10).

If a continuous curve were fitted to the segmented one, and its shape described by the equation,

$$Y = \frac{\ln (1+\mu x)}{\ln (1+\mu)}$$

the closest fit occurs when $\mu = 255$. This compares to $\mu = 100$ for D1.

The D2/D3 digital code for a sample is arranged as shown in Figure 11. Bit No. 1 of the 8 bits is used to designate the sign of the sample. If it is a positive sample, bit 1 is a 1. If it is a negative sample, bit 1 is a 0. Bits 2, 3, and 4 indicate on which of the eight segments of the compression curve the sample lies, after it is known whether the sample is positive or negative. Bits 5, 6, 7, and 8 tell which of the 16 steps within a segment most closely represents the height of a sample.

For example, if a small sample is to be encoded, it would be accomplished as shown in Figure 12. The sample at the coder input is positive, so bit 1, the sign bit, is a 1. Next, the sample falls within the coding range of seg-

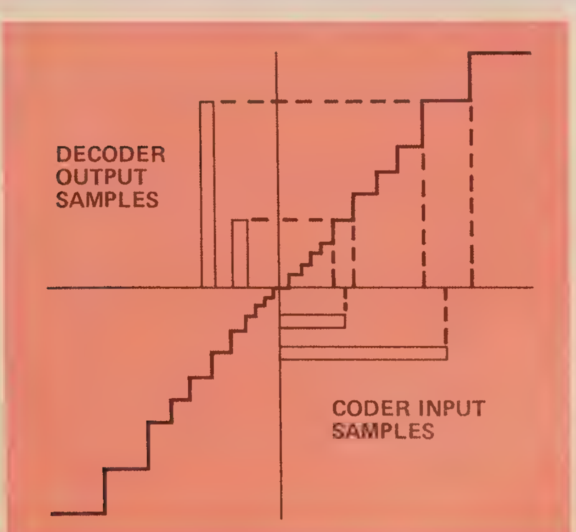


Figure 9. Non-uniform Codec transfer characteristic.

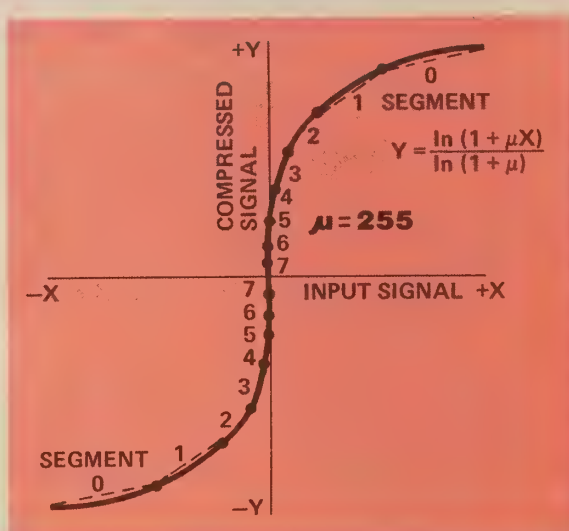


Figure 10. D2/D3 segmented compression characteristic.

BIT NO. 1 2 3 4 5 6 7 8
S A B C W X Y Z

S – SIGN BIT – 1 POSITIVE SAMPLES
– 0 NEGATIVE SAMPLES

ABC – DENOTES SEGMENT OF COMPRESSION CURVE USED IN CODING OF SAMPLE (8 BINARY COMBINATIONS)

WXYZ – DENOTES WHICH CODING STEP ON SEGMENT USED (16 BINARY COMBINATIONS)

Figure 11. D2/D3 compression code (8-bit).

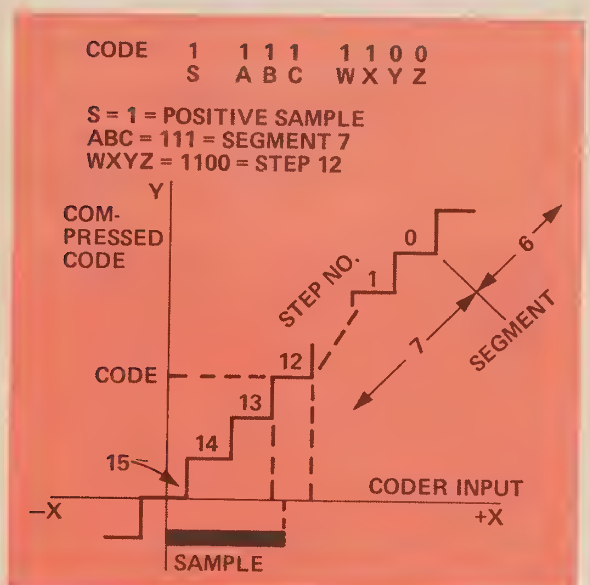


Figure 12. D2/D3 coding characteristic example.

ment 7, so bits 2, 3, and 4 are used to give the binary representation of 7, which is 111. Then, the final sample height is best described by step 12 of segment 7. Therefore, bits 5, 6, 7, and 8 are coded with the binary representation of 12, which is 1100. The 8-bit PCM word representing this sample is therefore: 1 1 1 1 1 1 0 0

The result of the 8-bit code and compression characteristic with μ = 255

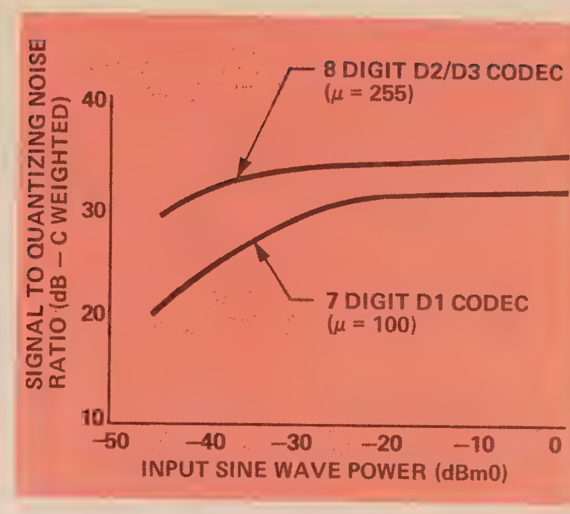


Figure 13. Signal-to-distortion performance (D1 and D2/D3 logarithmic codecs).

gives a substantially better signal-to-quantizing distortion ratio for the D2/D3 systems compared to that of the D1 system.

With 8 bits for coding vs 7 bits, signal-to-quantizing distortion ratio is improved by about 6 dB in the range of higher level signals. Then, the compression characteristic with μ = 255 gives the D2/D3 a relatively constant signal-to-quantizing distortion ratio over about a 40-dB range of input talker levels. This is an improvement over the μ = 100 compression characteristic for D1, which gives only about a 30-dB range over which the signal-to-quantizing distortion ratio is relatively constant (see Figure 13). This is one of the major improvements in the D2/D3 systems which was required to provide toll-quality service.

The following part of this series will deal with common equipment, channel-unit functions, and with some of the various signaling arrangements possible with modern PCM equipment.

BIBLIOGRAPHY

1. Crawforth, L.D. *The Increasing Flexibility of PCM Carrier.* San Carlos, California: GTE Lenkurt Inc., 1974.

2. de Leon, Jose C., ed. *Readings in Pulse Code Modulation.* San Carlos, California: GTE Lenkurt Inc., 1973.

PCM Update - Part 2

The signaling that has to be transmitted between two telephone offices takes two forms; Pulsing and Supervision. A PCM system is an ideal transmission medium for those signaling forms because it is a digital system transmitting digital information.

Part I of this series described some of the concepts and applications of PCM in subscriber carrier and TDM switching. This issue deals with PCM signaling arrangements, common equipment, and channel unit functions.

Telephone dial pulsing is relatively slow, compared to sampling rates in the PCM system. Therefore, the pulsing digits can be sampled, transmitted, and reconstructed with ease (see Figure 1). The reconstructed pulse is never perfect, but the distortion can be held as low as desired by adjusting the signaling sampling rate. In the D2/D3 carriers, this has been optimized to accommodate many different types of interoffice signaling.

Framing and Signaling Frame Identification

In the D2/D3 frame format, the 193rd bit – the timeshared or "S" bit – is reserved for terminal framing functions and signaling framing identification (see Figure 2). Terminal framing is required to align the receiver timing with the incoming bit stream from the remote transmitter so that channel word decoding can be accomplished in the proper sequence. To do this, terminal framing is sent as a unique repetitive pattern of 1 0 1 0 1 0---, and steps are taken in the transmitter to prevent any simulation of this pattern, on a continuing basis, by any voice or signaling transmission.

Since signaling is transmitted only once every sixth frame, a pattern is needed for the receiver to identify

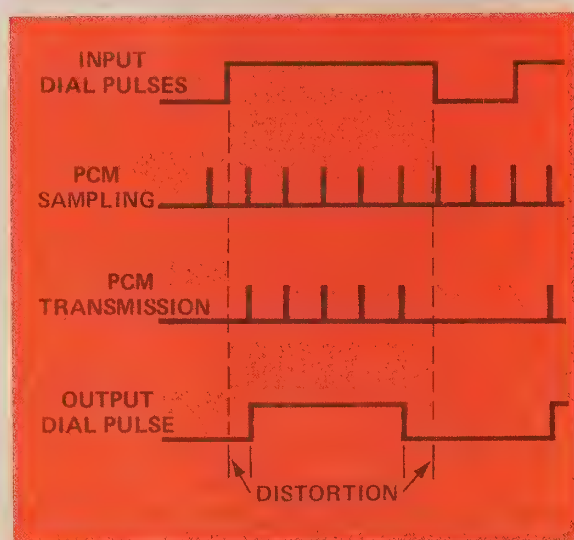


Figure 1. PCM signaling sampling.

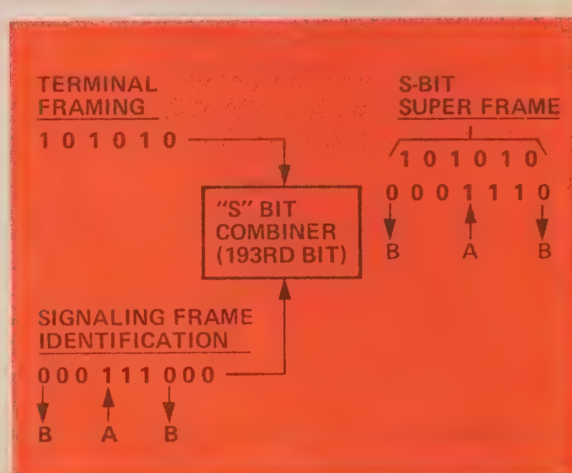


Figure 2. D2/D3 framing and signaling identification.

which of the six frames contains the signaling. This pattern is the signaling frame identification pattern, which is again a unique repetitive pattern of 0 0 0 1 1 1. These two patterns are combined to time-share the 193rd bit in each frame. A complete sequence of these time-shared bits requires 12 frames, and is called a "Super Frame."

With this signaling frame identification pattern, two signaling conditions, A and B, can be sent in alternating 6th frames. That is, conditions A and B are both sent within 12-frame increments, each condition appearing six frames apart from the other. These are then interleaved to send a signaling condition every 6th frame. This is accomplished in the following manner:

1. Signaling condition A is transmitted in the 8th bit of all 24 PCM channel words in the frame whose 193rd bit is a 1, preceded by the framing pattern 1 0 0 0 1.
2. Signaling condition B is transmitted in the 8th bit of all 24 PCM channel words in the frame whose 193rd bit is a 0, preceded by the framing pattern 0 1 1 1 0.

This framing scheme allows each voice channel to send up to two different conditions of signaling. This gives a capability of two signaling channels per voice channel, in each direction (see Figure 3). Since each signaling channel is sampled every 12th frame, this gives a sampling rate of 1.5 milliseconds (12 x 125 μs).

In the frames where signaling is allowed, each signaling channel uses the 8th bit of the PCM word. Since this 8th bit can be either 1 or 0, each signaling channel can send two states of signaling. Therefore, a total of four states of signaling can be sent in each direction of transmission.

These signaling channels can then be used either separately or combined, depending on the signaling requirements imposed by the service to which

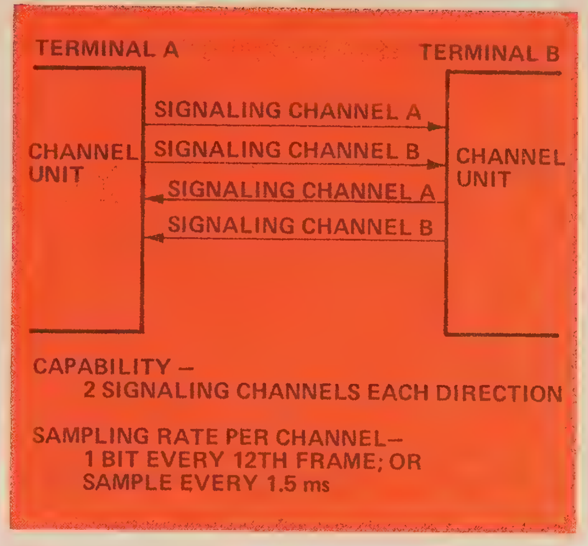


Figure 3. D2/D3 signaling channel capability.

the channel bank is committed. This feature is one of the greatest contributors to the flexibility of the D2/D3 PCM channel bank, allowing it to be used for signaling in subscriber, direct, toll connecting, tandem and intertoll applications merely by assigning these signaling channels in the appropriate way. For example, in E&M signaling, only two signaling states, on hook and off hook, are required in each direction (see Figure 4). This requires only one signaling channel; therefore, the two signaling channels available are combined to give a faster sampling rate of 0.75 milliseconds, or signal every 6th frame. Since low sampling distortion (less than 2%) is required in E&M

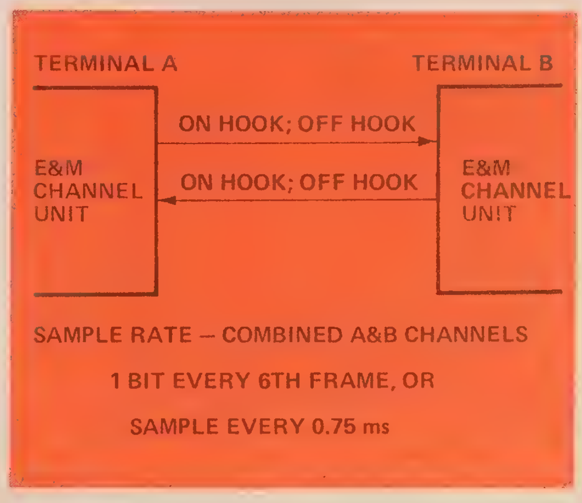


Figure 4. D2/D3 E&M signaling.

signaling, this rate of signaling transmission is required to accomplish it.

In contrast to E&M signaling, foreign exchange signaling requires up to four states of signaling in one direction. Also, two separate conditions, such as loop closure and ring ground, require transmission simultaneously (see Figure 5). Here, the two signaling channels can be used separately to send one condition every 12th frame; the signaling sampling rate of 1.5 milliseconds is adequate for the type of functions being transmitted.

Common Channel Interoffice Signaling

Use of the 8th bit in the PCM channel word for signaling in every 6th frame causes a degradation of about 2 dB in the signal-to-quantizing distortion ratio. Provision has been made in the D2/D3 format to upgrade the channel performance in the future by this 2 dB. This can be accomplished by allowing 8-bit voice encoding in all frames, including every 6th frame.

Signaling for all voice channels may be transmitted over a "common signaling channel" instead of using the 8th bit every sixth frame for each voice channel. First, all signaling for the voice channels is gathered into external equipment which multiplexes and codes it into a 4-Kbs data stream. This 4-Kbs data stream is substituted into the portion of the "S" bit stream previously occupied by the signaling frame identification pattern (see Figure 6). There is no frame now reserved for signaling, so the signaling frame identification pattern is not needed. If its share of the S bit data capacity is looked at as a 4-Kbs data stream, the common channel interoffice signaling (CCIS) data stream can be directly substituted. This flexibility of the D2/D3 format allows for future use of CCIS as it is introduced into newer type switching offices.

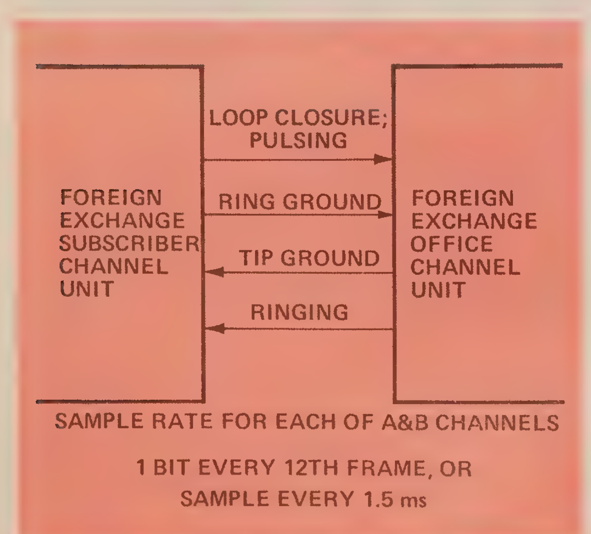


Figure 5. D2/D3 foreign exchange signaling.

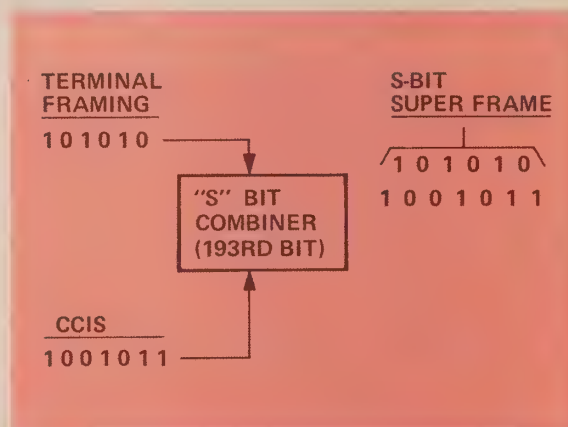


Figure 6. D2/D3 framing and common channel interoffice signaling.

Physical Organization of D2/D3 Carriers

In its simplest form, a PCM terminal contains the following functions:

(1) channel units to provide the individual voice channel interface between the PCM terminal and the telephone office, (2) a transmit common unit to provide all necessary functions to encode all voice channels for PCM transmission, (3) a receive common unit to recover the PCM transmission and decode it back to individual voice channels, and (4) an alarm and control unit to remove the terminal from service when a system failure occurs. If the interface between channel units

and common equipment is chosen carefully, both the simplest packaging of the terminal and the greatest flexibility can be achieved at the same time (see Figure 7). In all of the latest generation PCM terminals, this interface is chosen so that each channel unit contains all those functions which are truly limited to only that one voice channel; allowing common units to contain all those functions which are truly common to all voice channels in the terminal.

Common Equipment

The transmit common equipment (see Figure 8) contains a clock (at 1.544 MHz for D2/D3 terminals) and all the timing counters required for both common coding and timing functions and for individual channel sampling control. The coder converts channel PAM (pulse amplitude modulation) samples from a common interface bus to PCM words on a time multiplexed basis. Signaling is added to each channel PCM word at the appropriate bit time, and a bipolar converter converts from the unipolar pulses used in the terminal logic circuits to a bipolar signal more suitable for transmission on a repeatered line.

The receive common equipment (see Figure 9) starts the recovery by converting from the bipolar line transmission format to unipolar pulses which can be processed by logic circuits. A 1.544 Mbs clock is recovered from the incoming bit stream to synchronize the receive operations to the far-end transmit operations. Timing counters provide timing both for common decoding and for individual channel reconstruction of the transmitted signal. Signaling bits are recovered from the PCM stream and routed to the appropriate individual channels for signaling recovery.

Channel Units

The real flexibility of the PCM system is achieved in the channel units. The majority of the voice transmission functions are common to all types of channel units. By changing

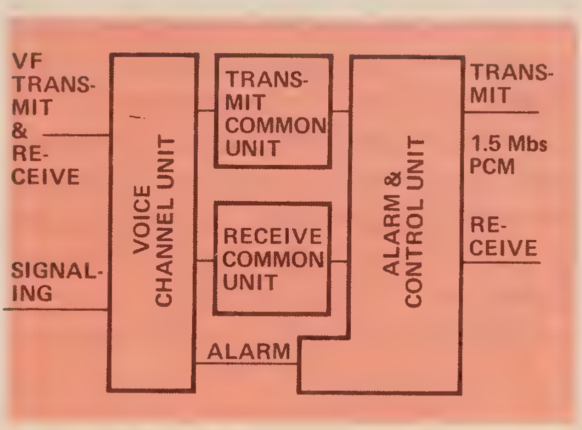


Figure 7. PCM terminal functions.

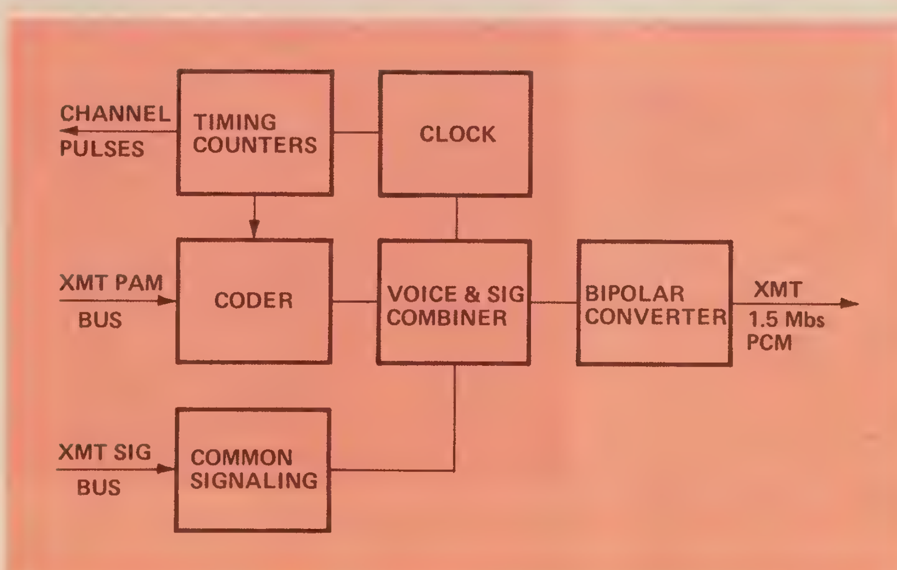


Figure 8. PCM common transmit functions.

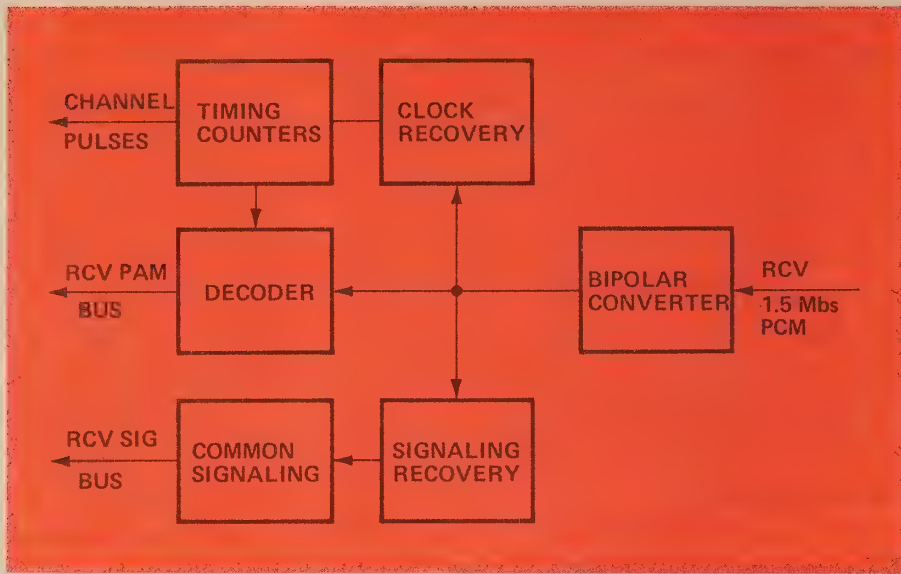


Figure 9. PCM common receive functions.

the signaling functions and office interfaces, many types of channel operations can be created, thereby creating many types of system applications. For example, in intertoll usage, the most common type of channel operation required is to provide 4-wire voice transmission and E&M —independent transmit and receive— signaling (see Figure 10).

In the transmit voice path, an amplifier provides level adjustment and office impedance matching. A low-pass filter limits the incoming frequency band to less than half the 8 kHz sampling rate, for proper sampling. Finally, a transmit gate connects samples of the vf signal to the common pulse amplitude modulated bus at the appropriate time to time multiplex all channel PAM samples.

In the receive voice path, a receive gate connects PAM samples from a common bus to the individual channel at the appropriate time to time demultiplex all channel PAM samples. The low-pass filter reconstructs the original vf signal, and the amplifier provides level control and office impedance matching.

In the signaling portion of the unit, a signaling transmitter converts from the signaling interface levels of the office, to logic levels usable in the terminal. Then a transmit signaling gate connects samples of the office pulsing or supervision to a common signaling bus at appropriate times to time multiplex all channel signaling samples. In the receive direction, a receive signaling gate connects samples from a common bus to the individual channel at the appropriate time to demultiplex all channel signaling samples. Then, a signaling receiver reconstructs the pulsing or supervision and provides the proper interface to the office.

To create another type of channel

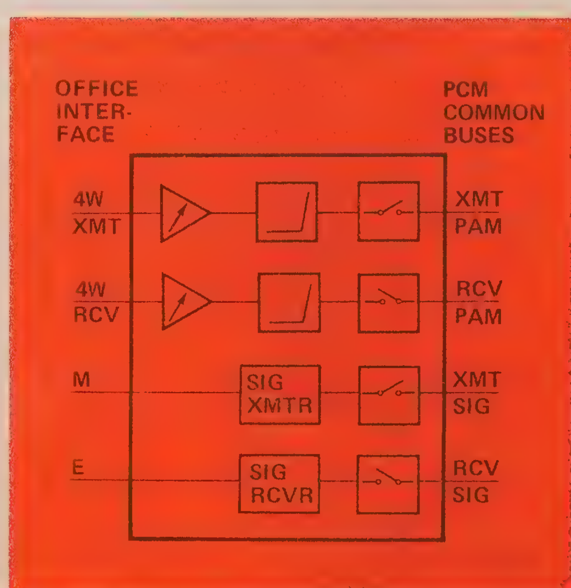


Figure 10. Four-wire PCM channel unit.

unit for different operations—direct trunk usage, for example—many of the channel functions stay the same. The requirements for different functions are only in the voice-path-office interface and in the signaling interface to the office. For instance, in a 2-wire loop signaling channel unit (see Figure 11), only a hybrid function is added to the amplifier, filter, and gate functions already available in the 4-wire design to provide for a different office interface in the voice path. Then, different signaling logic and office signaling interface is used with the signaling gate design already available to produce the required office interface. This same technique is also used to produce the many different channel units used in toll connecting and subscriber applications.

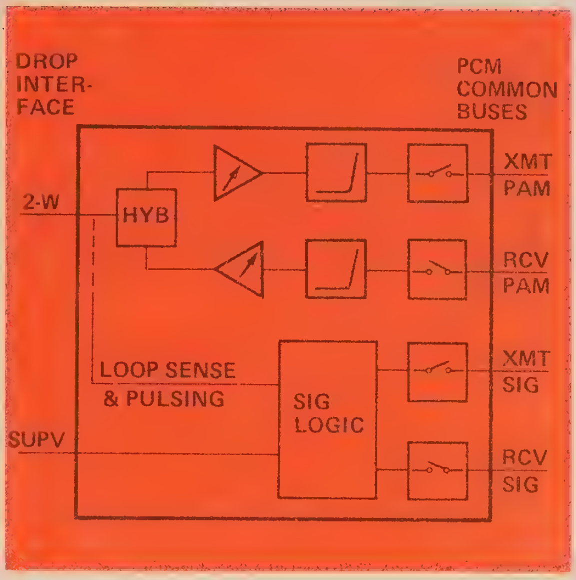


Figure 11. Two-wire PCM channel unit.

Important Factors

The principal factors contributing to the increasing flexibility of PCM carrier include a standardized D2/D3 PCM coding format which allows high enough voice quality for the most demanding portions of the telephone network, while allowing low enough cost to apply to the least stringent portions of the network, and time division multiplex digital switching machines to be applied at the network switching points. Additional factors that enhance the operation of PCM are the nature of PCM, which lends itself ideally to providing signaling and supervision transmission as a part of the PCM bit stream, and the physical organization of the latest generation D2/D3 channel banks, which allow standard interface PAM and signaling buses to provide maximum flexibility of application of signaling and voice functions.

BIBLIOGRAPHY

1. Crawforth, L.D. *The Increasing Flexibility of PCM Carrier.* San Carlos, California: GTE Lenkurt Inc., 1974.

GTE LENKURT

DEMODULATOR

SEPTEMBER 1974

PCM OVER MICROWAVE PART 1

The traditional approach to transmission of digital information has been with cable as the transmission medium, but the increased demand for digital microwave systems has led to a number of developments in the transmission of PCM over microwave radio.

The transmission of PCM (pulse code modulation) over microwave radio is particularly desirable when the geography of an area or some other obstacle impedes or prevents the accomplishment of cable plant. For example, PCM over microwave radio may be useful in situations where a body of water must be crossed to complete the communication path, where it is necessary to extend low-capacity trunks in rural areas, and where there is a requirement for temporary low-capacity trunks for special service or telephone restoration.

Recent developments at GTE Lenkurt have led to the introduction of equipment capable of multiplexing two T1-type PCM signals (each comprised of 24 PCM voice channels), and coding and shaping these in such a way that transmission over any frequency-modulated microwave radio is technically feasible. This type of equipment is thus capable of transmitting 48 PCM channels over one radio. By employing cross-polarized operation, which uses two radios on the same radio frequency (one radio using a vertically-polarized antenna system, the other a horizontally-polarized), this same equipment has the capability of utilizing the radio spectrum to transmit 96 PCM voice channels per radio channel.

An example of a hypothetical PCM-over-microwave system is shown in the simplified block diagram in Figure 1. The major components of the system are the digital multiplexer, the microwave radio baseband equipment, and the microwave radio transceiver. The input to the system is provided by two 24-channel D1, D2, or D3-type PCM terminals operating over a T1 line, or by any other type of equipment capable of producing a 1.544 Mb/s bipolar bit stream. The combination of any two 1.544 Mb/s channel banks can be used to meet subscriber, exchange, toll connecting, and intertoll applications, as well as low- or high-speed data transmission.

The Digital Multiplexer

The digital multiplexer combines two 1.544 Mb/s digital inputs into a 3.156 Mb/s signal, which is then converted to "modified duobinary" format for application to the radio baseband assembly. The modified duobinary output signal from the multiplexer is applied to the radio baseband equipment, where it is combined with voice-frequency order wire and alarm and continuity pilot signals. The composite signal is then level-coordinated and fed to the FM transmitter. The process is reversed in the receive portion of the system. Figure 2 shows the digital multiplexer in block diagram form. Low speed input and output signals to the multiplexer are designated DS1, which means "digital signal at the first level," and refers to the 1.544 Mb/s bipolar line signal from T1-type equipment such as channel banks, data terminals, and repeatered lines.

The transmit portion of the multiplexer synchronizes and interleaves the two asynchronous, bipolar input signals into a single bit stream, which is then scrambled, encoded, and converted to a modified duobinary waveform for application to the baseband equipment.

The receive portion of the multiplexer accepts a signal from the radio baseband assembly. It filters the received signal, blocks the radio pilot frequency, and decodes and descrambles the signal. The signal is then demultiplexed and desynchronized into two separate 1.544 Mb/s streams. Two new bit streams are then produced that are identical to those applied at the transmit end.

Synchronization

The bit rate of the asynchronous T1 lines coming into the multiplexer is a nominal 1.544 Mb/s; that is, each T1 line may have a bit rate that is sometimes higher, and sometimes lower, than the desired 1.544 Mb/s. This results in a continual shift in phase between the two T1 lines, which prevents the direct interleaving of bit streams into a single bit stream at twice the frequency. In addition, instantaneous frequency changes (timing jitter) that may occur in an individual T1 bit stream also prevent synchronization of the two T1 bit streams. To handle both of these problems, an eight-cell elastic store is included in the synchronizer portion of the "syndes" (synchronizer-desynchronizer) unit (see Figure 3). Each syndes extracts a clock (timing pulses) from the incoming data, and slices the data to convert it from a bipolar format to unipolar format (ones and zeros).

In the synchronizing process, after the DS1 streams have been converted to unipolar format, dummy or "stuffed" pulses are inserted, where needed, into each incoming T1 pulse stream, thus bringing the bit rate of each up to some common value before multiplexing. Each T1 pulse stream is written into the elastic store by means of the write clock derived from the incoming T1 pulse stream. Taking a simplified example, if the DS1 data coming into the synchronizer is at 1.544 Mb/s, the clock or timing pulses that are extracted from the data will be at 1.544 MHz. A synchronous T1 pulse stream is read out of the elastic store by means of a local 1.5458-MHz clock that is common to both synchronizers. The common clock rate of 1.548 MHz is a critical frequency, since too high a frequency would cause bits to be read before they actually appeared, and too low a frequency would result in an occasional loss of a bit. Since data must be written into the elastic store before it can be read, the read clock must operate at a faster rate than the write clock (1.544 MHz).

The read clock cannot be allowed to overtake the write clock to the extent that it attempts to read a bit that isn't yet present in the store. (This event is called "spilling" the elastic store.) To prevent this occurrence, the phase comparator (which compares the phase of the write clock with that of the read clock), through associated control circuitry, slows up the read clock by inhibiting readout for one time slot, and stuffs a dummy pulse into the synchronous pulse stream. This maintains the proper output bit rate and allows the write clock to again precede the read clock, thus keeping the elastic store at least partially full. Any bits inserted by stuffing are removed during the demultiplexing operation.

The Transmit Common Unit

The transmit common unit provides pulse stuffing for both syndes. Each stuffed bit is identified by a complex framing scheme which allows recogni-

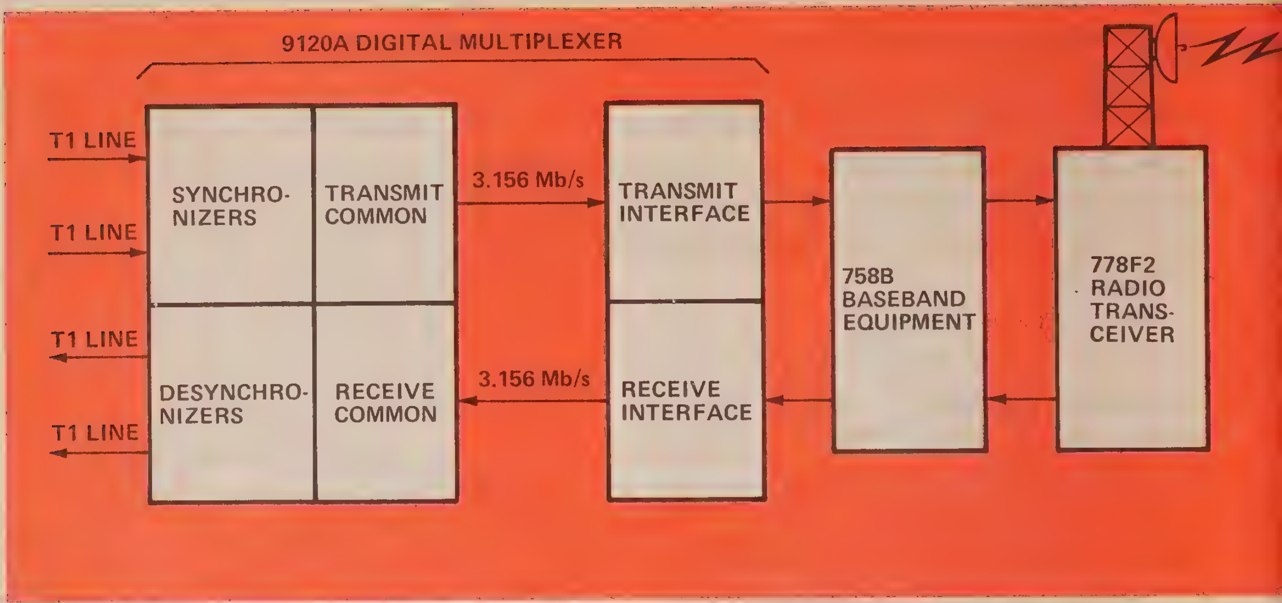


Figure 1. Block diagram of a PCM-over-microwave system with two 1.544-Mb/s inputs.

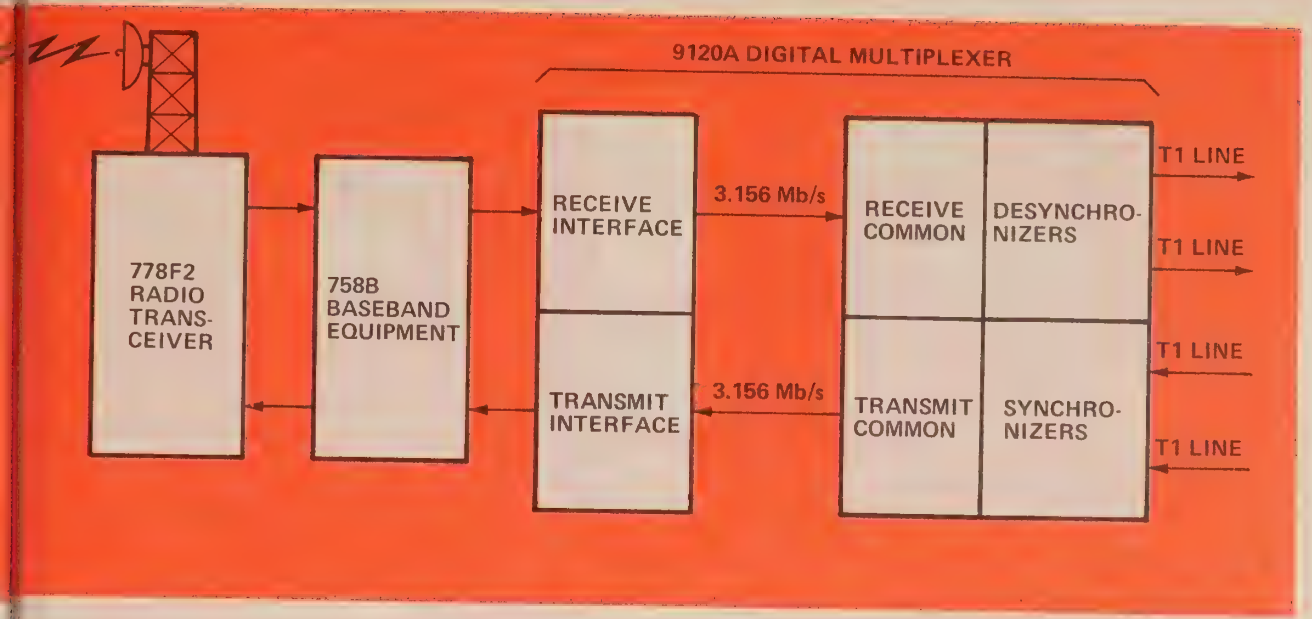
tion of the stuffed bits at the receive end so that they may be eliminated from the bit stream. The transmit common unit provides control bits (including framing bits) as well as stuffing bits. The framing bits, which are added to the data stream at precise intervals, are actually a form of bookkeeping: they keep track of which bits belong to T1 line A, and which to T1 line B.

The output from the syndes is two bit streams of ones and zeros at the same clock rate (1.5458 MHz). The transmit common unit interleaves the two bit streams, adds control bits, and clocks out a bit stream of 3.156 Mb/s; this is slightly more than twice the bit stream frequency (1.544 Mb/s) due to the addition of control bits and stuffed bits.

Demultiplexing

The receive common unit performs the opposite functions of the transmit common unit. The 3.156-Mb/s signal enters the receive common unit, where the framing bits are removed and monitored. The overall condition of the transmission medium may be determined by monitoring the framing bits. The information bits constitute totally random data, but the insertion of framing bits guarantees a point of reference at a specific location. If the framing bits are repeatedly not in the prescribed location, synchronization is momentarily lost, and the multiplexer initiates a resynchronizing process.

The receive common unit monitors the incoming 3.156-Mb/s signal, examining incoming control bits for a predetermined pattern. The receive common circuitry removes the stuffing and control bits, then splits the remaining bit stream into two 1.544-Mb/s streams. Each bit stream is then applied to the desynchronizer portion of the appropriate syndes (see Figure 3), where it is retimed with a phase-locked loop through another 8-cell elastic store to eliminate the gaps caused by the bits removed at the receive end. At this point the bit streams are again asynchronous, and the original data at the input of the multiplexer is reproduced at the output of the demultiplexer.



To review, two asynchronous (different frequency) bit streams have been interleaved into a single bit stream at slightly more than twice the nominal bit rate of 1.544 Mb/s. Pulse-stuffing synchronization is used to bring the bit rate up to 3.156 Mb/s for output to the transmit interface unit. At the far end, just the opposite process takes place to reproduce the data.

Transmit Interface Unit

The transmit interface unit converts the binary (ones and zeros) 3.156-Mb/s signal to one suitable for application to the baseband portion of the radio. If the binary signal were to be applied directly to the radio transmitter, the resulting radio signal would occupy more than the maximum allowable bandwidth (3.156 MHz) for a 2-GHz system. One major objective for the interface unit is thus to compress the incoming data so that the radio signal will not exceed 3.156-MHz of bandwidth. Careful system design is necessary to meet this objective, since the compression of data to fit within the required bandwidth must be accomplished without an objectionable increase in error rate during deep

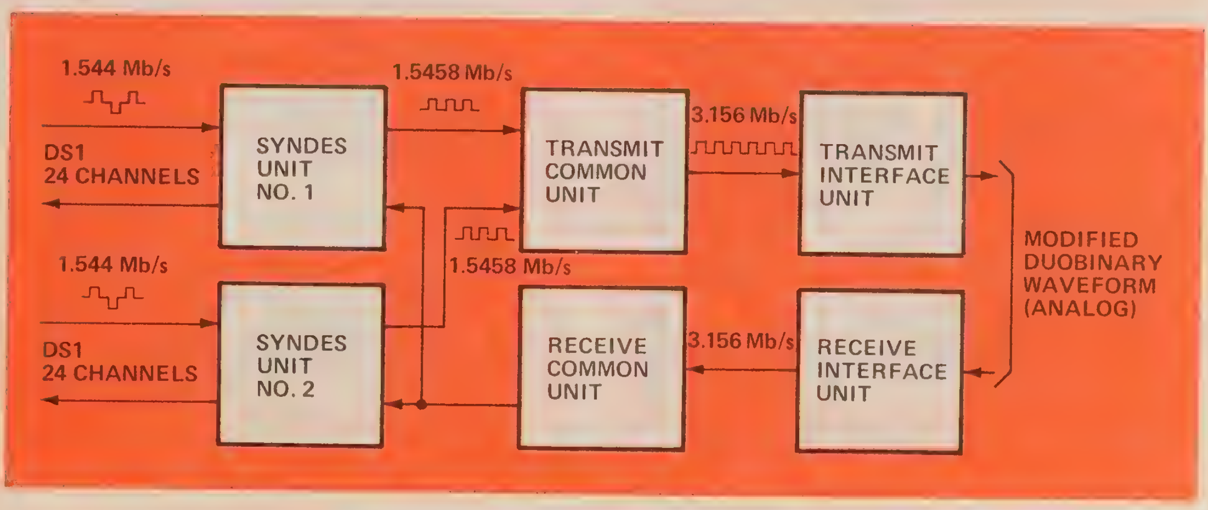


Figure 2. The digital multiplexer processes two DS1 inputs for application to microwave baseband equipment.

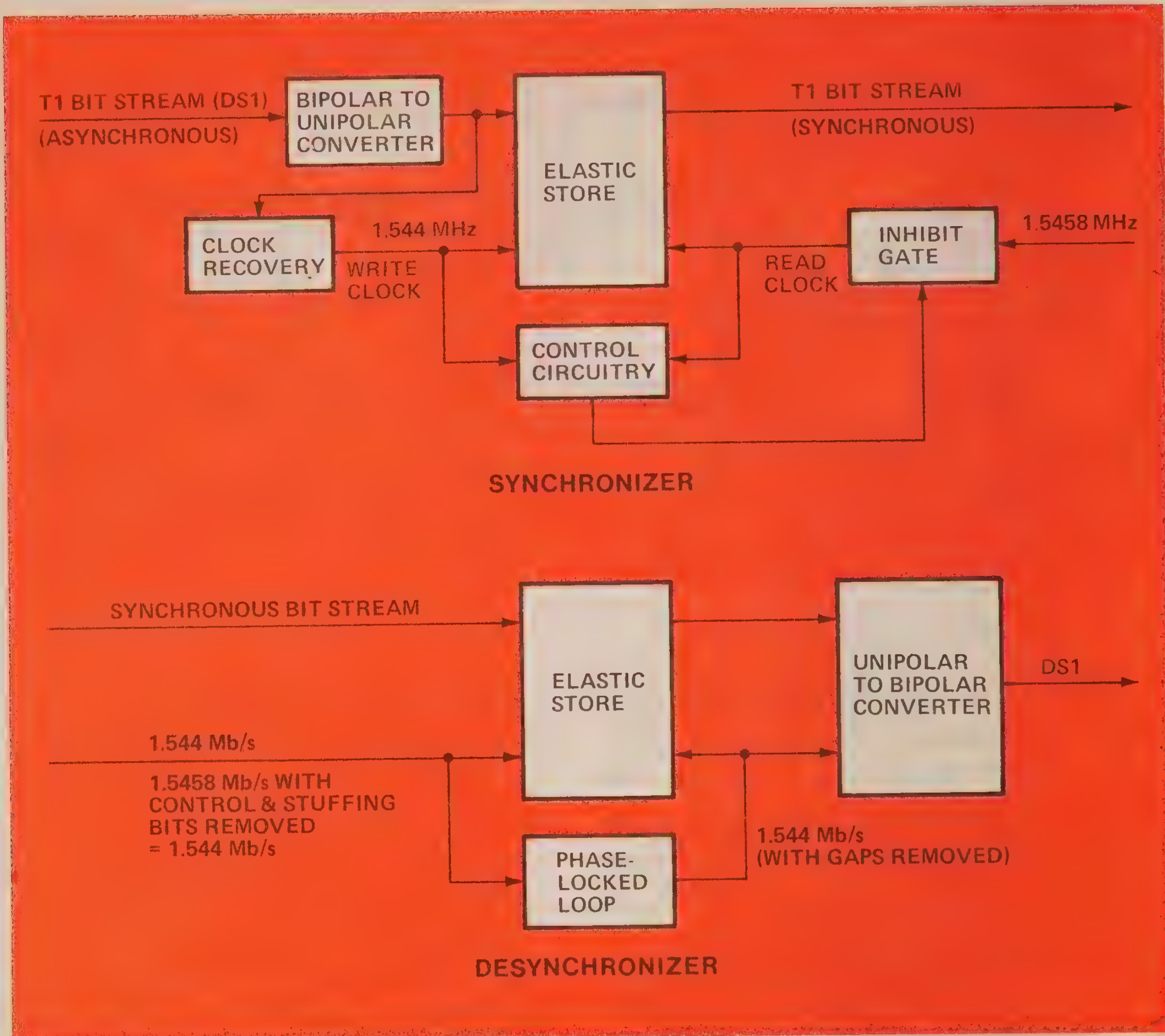


Figure 3. Synchronization and desynchronization take place in the syndes portion of the digital multiplexer.

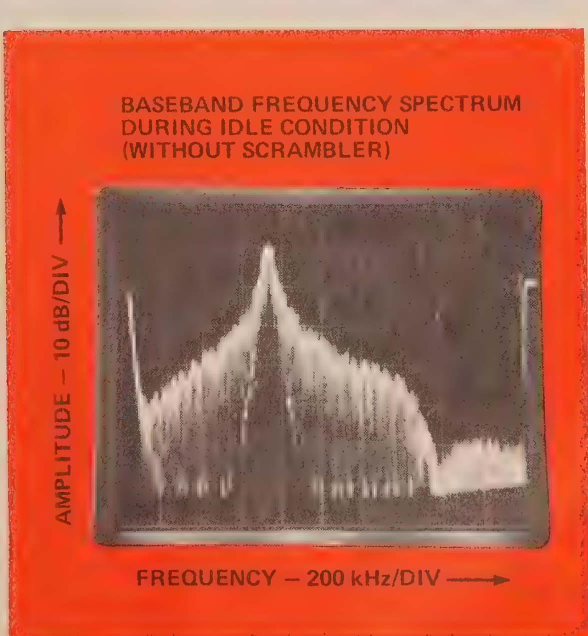


Figure 4. Scrambling is necessary during idle conditions to prevent the occurrence of power concentration at discrete frequencies.

atmospheric fades. The transmit interface unit also removes dc and low-frequency components so that an order-wire signal can be transmitted with the main body of data.

Scrambling

The sequence of events which takes place in the transmit interface unit after the two incoming DS1 signals are converted into a single 3.156 Mb/s stream is scrambling, conversion to modified duobinary encoding waveform, and filtering. There are two inputs to the transmit interface: a clock from the transmit common unit at 3.156 MHz, and data from the transmit common unit at 3.156 Mb/s. The main reason for scrambling the binary signal is to eliminate discrete frequency components during idle conditions. An idle condition may exist, for example, during the very early hours of the morning when perhaps all 48 channels are not in use, except for the transmission of framing bits. During this idle period, repetitive frequency patterns constituting a heavy concentration of system power at particular frequencies are likely to occur; such power concentrations can cause interference in other radio channels in the vicinity. Scrambling eliminates this condition by creating a continuum of power over all the channels in the system during idle conditions. The even distribution of available power over the entire allocated frequency spectrum reduces interference to nearby radio channels. Figure 4 shows a comparison of idle-condition output spectrums with and without a scrambler in the system. The scrambling process consists of combining the data entering the transmit interface unit with a pseudo-random bit sequence generated by a conventional feedback shift register.

An important consideration in the design of an overall transmission is the inclusion of a voice-frequency order wire in the radio baseband equipment; how this is achieved, while conforming to certain bandwidth requirements, will be the starting topic of the next Demodulator.

BIBLIOGRAPHY

1. Cheng, C.S., "The 9120A Digital Multiplexer (SSS-9120A)," San Carlos, CA: GTE Lenkurt Inc., (March 1974).

2. Crawforth, L.D. and J.D. Olson. "Maximum T1 Repeatered Line Length Considerations," *International Conference on Communications*, Vol. II, (June 1973), 32-28 – 32-33.

3. Karchevski, R.A., "The 91965 Syndes Unit (PSS 91965-M1)," San Carlos, CA: GTE Lenkurt Inc., (March 1974).

4. Kline, R.L., "The 91963 Alarm Unit (PSS 91963-M1)," San Carlos, CA: GTE Lenkurt Inc., (March 1974).

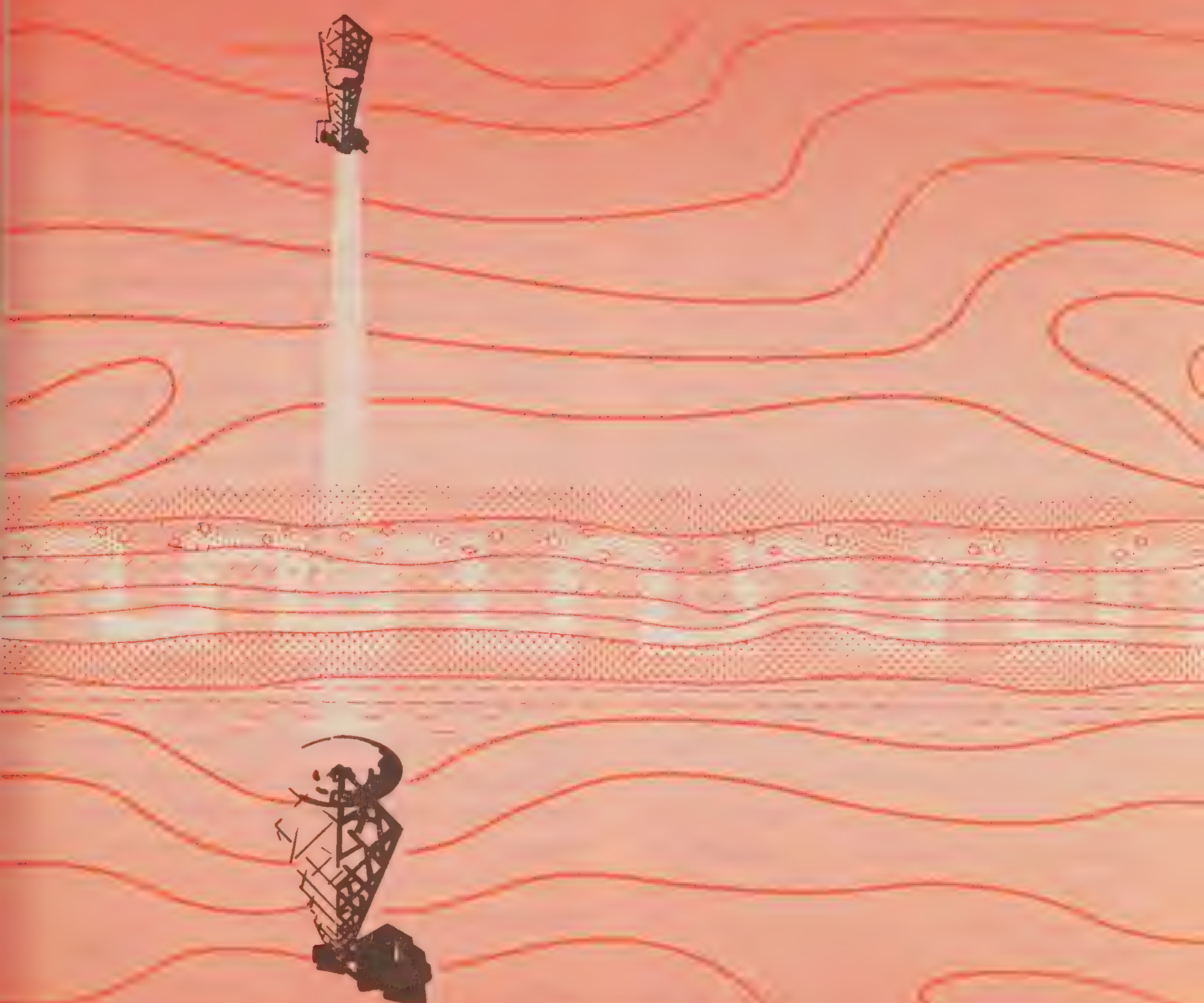
5. Olson, J.D., "A Method of PCM Transmission Over 2-GHz Microwave Radio," *International Conference on Communications*, (June 1974), 33C-1 – 33C-5.

6. Swartz, L.T., "Performance Analysis of a Three-Level Modified Duobinary Digital FM Microwave Radio System," *International Conference on Communications*, (June 1974), 5D-1 – 5D-4.

GTE LENKURT

DEMODULATOR

OCTOBER 1974



PCM OVER MICROWAVE PART 3

"In digital communication systems, speech, print, pictures, and computer data are all represented by binary signals. These signals have inevitably to be transmitted over band-limited communication channels, and the bandwidth is usually at a premium. To conserve bandwidth, it is usually necessary to convert the binary signals into other digital signals that require less bandwidth." A. Lender

The September, 1974, issue of the *Demodulator* began a discussion of the transmission of PCM signals over microwave radio. Covered in that issue were some of the functions performed by various components in the digital multiplexer, such as multiplexing, synchronization, desynchronization, and the scrambling of the binary signal to eliminate discrete frequency components during idle conditions. This issue discusses the importance of bandwidth conservation, and the methods by which it is accomplished.

One method of achieving bandwidth conservation is by compressing the incoming digital source information, which is the equivalent of removing redundancy from a message. Human speech, for example, has a redundancy factor of nearly 50%; but, while speech can be compressed, the cost of doing so presently excludes the possibility of commercial utilization. Another technique for the conservation of bandwidth is transformation of the binary source signal into a digital signal requiring less transmission bandwidth.

Duobinary Technique

The transmission of digital data over microwave radio requires that the bandwidth of the data be limited to comply with certain requirements. An approach taken in the past to reduce the bandwidth of the digital bit stream is duobinary encoding, a type of correlative encoding technique invented by Dr. Adam Lender of GTE Lenkurt. (See the February, 1963, issue of the *Demodulator* for a discussion of the duobinary process.) The duobinary power spectral density appears as shown in the example in Figure 1. The energy is contained between zero and 1/2T Hz (where 1/T is the bit rate in bits per second and T is the duration of one bit in seconds) by filtering techniques which eliminate any higher frequencies.

The frequency of the duobinary power spectrum beyond which there is no energy is 1/2T Hz. This cuts off the signal at half its bit rate, meaning, that of the original spectral density of the input binary pulse train, only frequency components of up to 1/2T Hz are being allowed through the system. The straight duobinary spectral density has most of its energy at dc and low frequencies, a situation which poses certain problems when designing voice-frequency order wire for a microwave system.

Order Wire

An important consideration in any microwave transmission system is the inclusion of a voice-frequency order wire in the radio baseband equipment. Order wire is essentially a telephone channel between communications sites, which maintenance personnel can use to resolve problems in the system. The order wire can be implemented in either digital (as part of the bit stream) or analog form. One of the major objectives in implementing order wire in the GTE Lenkurt microwave system is to be able to use

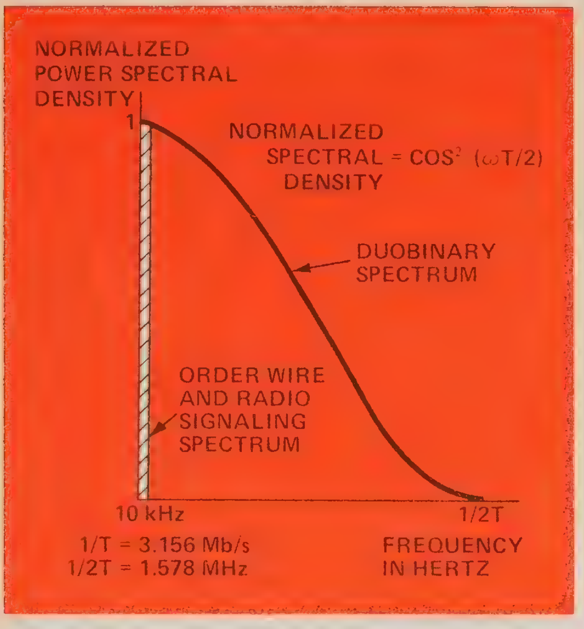


Figure 1. The conflicting spectrums of order wire and duobinary require modification of the duobinary waveform.

conventional, unmodulated analog order wire such as is used in an analog microwave system.

One possible method of providing order wire and radio signaling information is to place these channels at the higher end of the spectrum, above the digital information. This method, however, is inconvenient, since additional circuitry is required to translate the voice frequencies to the upper end of the spectrum, and the noise on the microwave radio system is greater in that part of the spectrum due to the so-called triangular noise voltage distribution. Also, bandwidth restrictions make it necessary to use a coding technique which affords more efficient use of the bandwidth than this method allows. A more practical approach is to place the order wire and service channels at the low end of the energy spectrum; in this way, no additional modulation of the voice signals is required, and bandwidth is conserved.

The basic design of the GTE Lenkurt PCM microwave system requires that both the voice-frequency order wire spectrum and that of the 48 PCM channels be transmitted over the same bandwidth. To accomplish this, however, the distribution of the baseband energy spectrum must be modified in such a way that voice-frequency order wire and signaling channel information do not interfere with the digital information. Clearly, combined transmission of the two spectrums (see Figure 1) would result in mutual interference, since both have significant amounts of power at the low-frequency end.

To accommodate the voice frequencies and signaling at the low-frequency end, the coding scheme should result in a frequency spectrum with a minimum of power concentrated at the low end. Also, to conserve bandwidth, some form of correlative encoding (correlation between digits that are a certain number of time slots apart) is desirable. The method used by GTE Lenkurt to transmit order wire and multiplexed digital information simultaneously within the same frequency spectrum differs somewhat from the straight duobinary technique, and is termed Modified Duobinary Coding.

Modified Duobinary Technique

The modified duobinary approach to digital transmission (also invented by Dr. Lender) has two important advantages: conservation of bandwidth, and allowance for an economical order wire system within the required bandwidth. Modified duobinary encoding compresses the bandwidth of the binary data through a series of filtering processes, which ultimately yield the power spectrum shown in Figure 2.

In the GTE Lenkurt PCM microwave system, the frequency band between 0 and 10 kHz in the modified duobinary spectrum is reserved for order wire and radio signaling information. Like duobinary, modified duobinary still cuts off at 1/2T Hz; unlike duobinary, modified duobinary has a

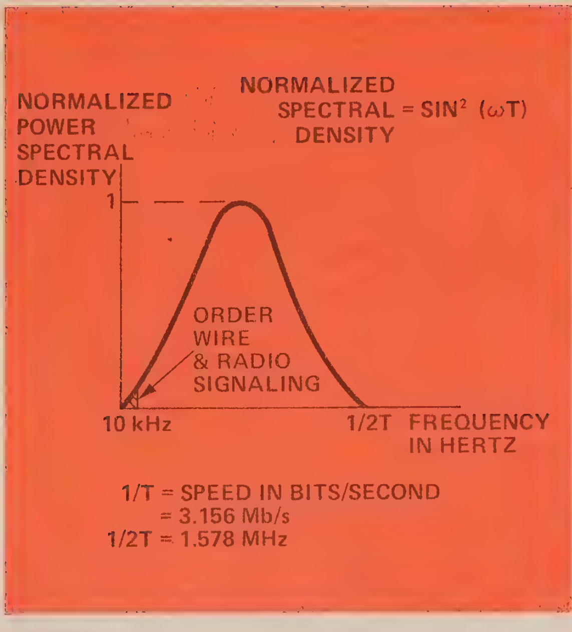


Figure 2. The modified duobinary spectrum contains negligible power at the low-frequency end.

negligible amount of energy at the low-frequency end. The power is contained between zero and 1.578 MHz by filtering techniques which eliminate any higher frequencies. In this case, the 1.578 MHz is exactly half of the incoming bit rate of 3.156 Mb/s, which corresponds to that of a GTE Lenkurt system designed for the transmission of 48 PCM voice channels over microwave radio.

Part of the modified duobinary process takes place in the transmit interface unit (see Figure 3) of the digital multiplexer. As discussed in the previous issue (PCM Over Microwave, Part 1), the scrambler prevents discrete frequency tones from being transmitted by the radio. The modified duobinary encoder encodes the 3.156-Mb/s binary data for the modified duobinary processing. Two additional steps are required for conversion to a modified duobinary waveform. In the first step, a three-level signal is formed from the encoded binary signal by the three-level converter. The second step is an analog function that is divided between the transmitting and receiving filters.

The conversion to a three-level signal constitutes the initial conversion step to a modified duobinary signal. The transmit filter and delay equalizer perform partial conversion of the three-level signal into the modified duobinary format. The remaining conversion is done at the receive filter. The delay equalizer minimizes the differential delay of the transmit and receive filters. If, for example, all the filtering were to be done at the transmitter, during a multipath fade, where increased noise is introduced into the receiver, there would be no further filtering available, and error performance would be degraded. Conversely, it would be undesirable to perform all

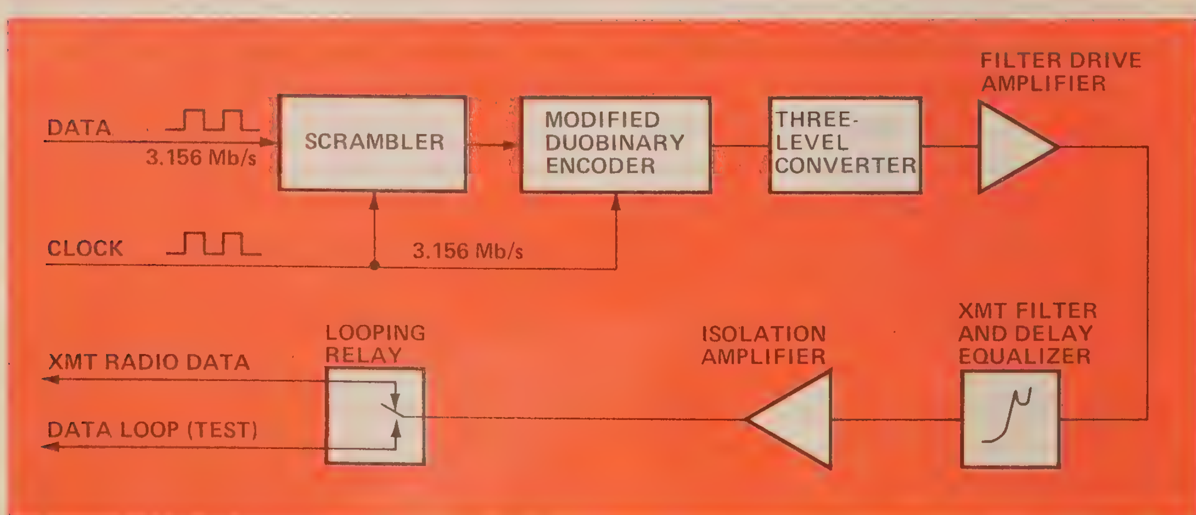


Figure 3. The first part of the conversion to a modified duobinary waveform takes place in the transmit interface unit.

the filtering in the receiver, since this would mean that a wideband signal would have to be transmitted, resulting in excessive use of the frequency spectrum. The solution to this problem is to do partial filtering at the transmitter, and the rest at the receiver. Full conversion to modified duobinary waveform does not actually take place until the output of the receive filter; a complete modified duobinary signal is therefore not transmitted through the medium. And, it is only at the output of the receive filter, in the receive interface unit, that the modified duobinary eye pattern may be observed.

The Eye Pattern

The eye pattern of a random pulse train is formed by dividing the pulse sequence into n-bit segments and superimposing the segments over the n-bit interval. By connecting the pulse stream to an oscilloscope that is synchronized externally with the clock that drives the pulse train, a pattern resembling a human eye is formed. The eye pattern may be produced to monitor overall system performance. By observing the eye, distortion and intersymbol interference can be measured in terms of vertical and horizontal eye openings relative to an undistorted eye. Figure 4 shows a typical eye pattern for a modified duobinary signal.

The eye pattern is sampled by the clock at some optimum point, which will yield either a –1, 0, or +1. The performance of the system depends on how well it can extract a signal in the presence of noise. The theoretical rms signal power-to-rms noise power ratio required for a bit error rate of 10^{-6} (one error every million bits) using the duobinary technique is approximately 17 dB. This type of noise performance, however, is theoretical, and would apply only to a perfect system; practically, a typical duobinary system has a noise performance of between 18 dB and 20 dB.

The modified duobinary signal is more susceptible to intersymbol interference than is a straight duobinary waveform. Unlike the duobinary format, which stipulates that a transition from +1 directly to –1 cannot take place without a transition to the zero level first, the modified duobinary format allows direct transitions from the positive to the negative level, thus increasing intersymbol interference. In effect, the more intersymbol interference, the more accurate the clock must be. Theoretically, there is a 2-dB penalty in the transition from duobinary to modified duobinary, with the expected signal-to-noise ratio for modified duobinary being around 20 to 22 dB.

Clock Recovery

The receive interface unit is shown in Figure 5 in block diagram form. The receive filter converts the signal into the modified duobinary waveform and limits out-of-band noise generated in the radio receiver. The slicers convert the analog signal into two binary bit streams. Each time the clock pulse rises, it is performing the function of "looking" at the information at a

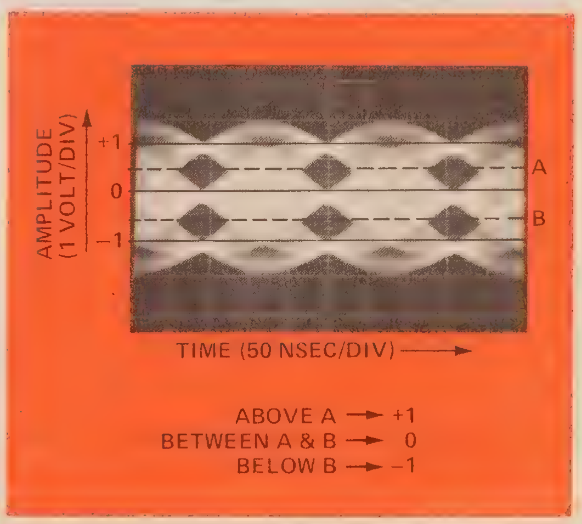


Figure 4. The eye pattern reveals overall system performance.

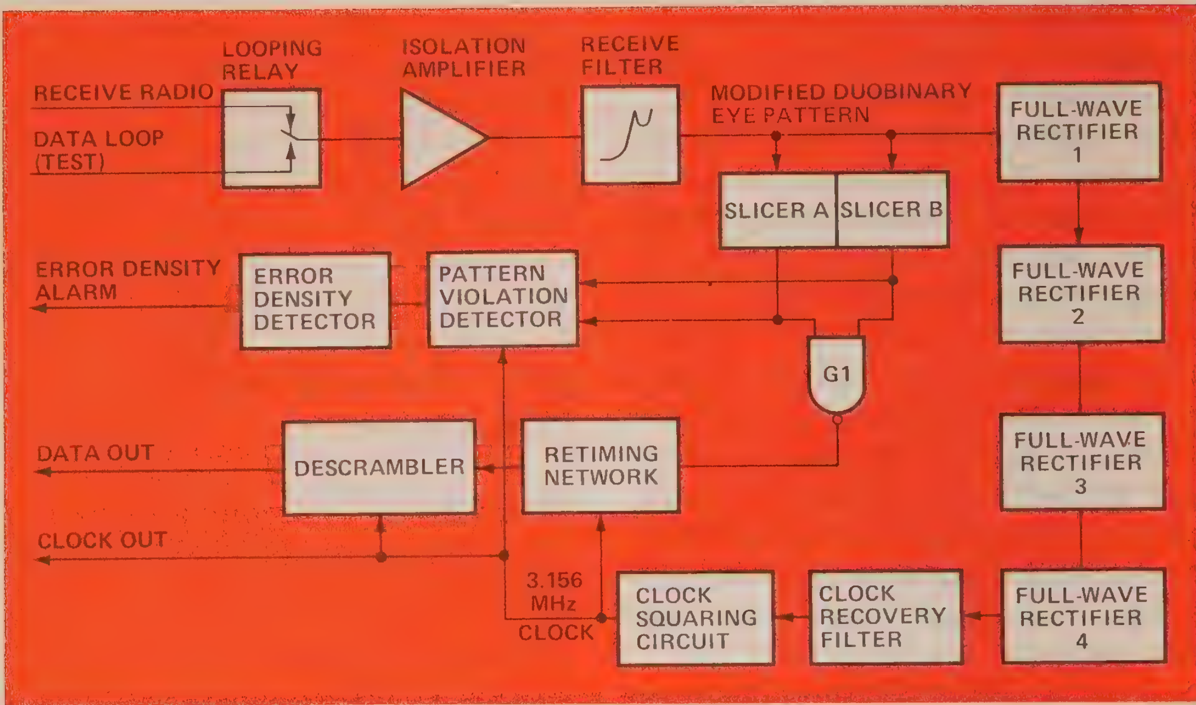


Figure 5. The receive interface unit completes the conversion to modified duobinary, recovers the clock, and descrambles the incoming signal.

particular point in time. The clock must look directly within the area represented by the eye, since that is where the desired information is contained. The rising edge of the clock pulse must therefore sample as close to the center of the eye as possible. Essentially, the clock must decipher whether the information at one particular instant is −1, 0, or +1, at the output of the slicer. Four full-wave rectifier circuits connected in tandem convert the incoming analog signal into a signal with a strong 3.156-MHz component. The clock recovery filter allows only the 3.156-MHz component to pass, thus producing a 3.156-MHz sine wave, which the clock-squaring circuit converts into a 3.156-MHz clock. The descrambler, acting to complement the scrambler in the transmit interface unit, descrambles the data into the original binary data. The pattern violation detector detects modified duobinary pattern violations, which result in errors. The error density detector triggers an alarm when the error rate exceeds 10^{-5} (one error in 100,000 bits).

The overall clock recovery process is especially interesting in this system. Specifically, there are two design objectives which the clock recovery circuit must meet: (1) To produce a clock with a minimum amount of phase jitter, and (2) to minimize clock phase shift caused by temperature variations and variations in the transmit clock frequency. The overall clock recovery process consists of producing a signal derived from the modified duobinary eye pattern with a strong 3.156-MHz component, and filtering the signal to select only that component.

Due to the spectral content of the modified duobinary signal, four successive full-wave rectifications produce a signal with a very strong 3.156-MHz component suitable for filtering. Each of the four full-wave rectifier circuits consists of a constant current source with temperature compensation and a differential amplifier with a gain of 2. The gain of 2 is provided in each rectifier so that the peak-to-peak signal amplitude is not reduced after each rectification.

After four successive full-wave rectifications, the signal frequency spectrum yields a sharp peak at the bit frequency of 3.156 MHz, along with an indefinite number of weaker out-of-band components. This signal is then passed on to the clock recovery filter, to assume the function of the clock.

To meet the first design objective of minimum phase jitter, the clock recovery filter must have a high enough Q so that the out-of-band frequency components are significantly rejected. However, to meet the second design objective of minimum clock phase shift, the Q must not be so high as to produce a phase characteristic which varies greatly with frequency. To meet both objectives satisfactorily, a GTE Lenkurt crystal filter was designed with a Q of approximately 1000 (a low Q for a crystal filter), and with a gradual phase characteristic, thus achieving a combination of minimum jitter and gradual phase shift. A graph of amplitude and phase shift versus frequency is shown in Figure 6. In the band of interest, the gradual phase characteristic of 0.03 degree per Hertz insures a phase-stable clock in spite of small offsets in the transmitting clock. Specifically, system design allows the transmit clock to vary by no more than ±100 hertz, thus preventing the receive clock from varying

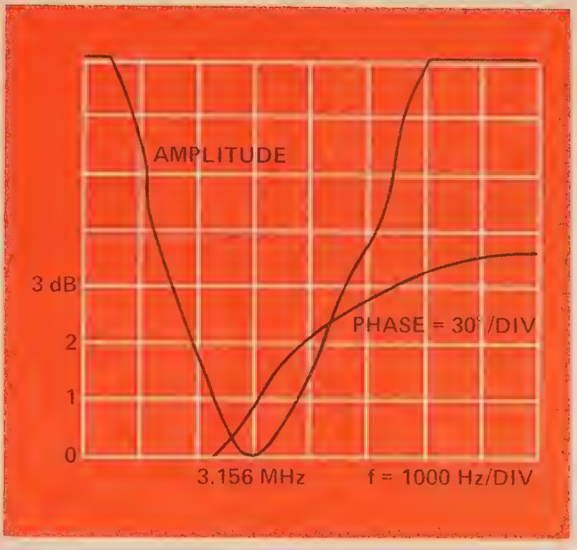


Figure 6. Amplitude and phase shift versus frequency.

more than ±3 degrees. Less than 10 nanoseconds of peak-to-peak jitter appear on the clock, with an increase to 20 nanoseconds at a signal-to-noise ratio of 20 dB. The filter is stable with temperature, having a 0.22 degree/°C phase shift, and a phase shift close to 0.03 degree per Hertz due to frequency offsets in the transmit clock.

This series of articles, up to the present, has dealt with some of the more significant points pertaining to the operation of specific system components involved in the transmission of PCM over microwave radio. The following issue (Part 3) will be a general discussion of the overall operation of the PCM-over-microwave system.

BIBLIOGRAPHY

1. Lender, A. "Correlative Level Coding for Binary-Data Transmission," *IEEE Spectrum*, Vol. 3, No. 2 (February 1966).

2. Lender, A. *Correlative Signal Processing*. Stanford, CA: Stanford University for Systems Research, 1972.

3. Olson, J.D. "A Method of PCM Transmission Over 2-GHz Microwave Radio," *International Conference on Communications* (June 1974), 33C-1 – 33C-5.

4. Swartz, L.T. "Performance Analysis of a Three-Level Modified Duobinary Digital FM Microwave Radio System," *International Conference on Communications* (June 1974), 5D-1 – 5D-4.

GTE LENKURT

DEMODULATOR

NOVEMBER 1974

PCM OVER MICROWAVE PART 3

The main impairments to good overall performance of a PCM-over-microwave system are far different from those that degrade either PCM cable carrier or FDM-over-microwave systems.

The first two parts of this series on the transmission of PCM over microwave radio have dealt mainly with the internal operation of a digital multiplexer (the GTE Lenkurt 9120A), which processes the PCM information into an analog signal suitable for transmission over a standard FM microwave radio system. This issue discusses the overall performance of a PCM-over-microwave system, using as a hypothetical example, a low density, 2-GHz system intended for short haul local exchange service.

The performance of a PCM cable system may be degraded by such nongausian electrical events as office noise, lightning hits, switching noise, and crosstalk between cable pairs. FDM-over-microwave is affected by delay and linearity inequalities within the radio system, by thermal or fluctuation noise, by path propagation fades to the noise threshold or mute point of the radio receiver, and by low levels of co-channel rf interference. PCM-over-microwave system performance is influenced primarily by multipath propagation fades that introduce excessive hits (errors) into the PCM bit stream, and by high levels of co-channel rf interference.

System Performance Measurement

The most meaningful measure of performance in a digital system is the bit error rate (BER) at the receiver digital demultiplexer output. The bit error rate is given as the number of bits in error, divided by the total number of bits sent. For example, a BER of 10^{-6} means that one error occurs in every one million transmitted bits.

The bit error rate depends primarily on the value of the signal-to-noise (S/N) ratio at the receiver. The S/N ratio is the difference between signal and noise strength in dB. A S/N ratio of 10 dB, for example, means that the rms value of the signal is 10 dB greater than the rms value of the thermal noise. System performance is determined by measuring signal-to-noise ratio at a particular bit error rate. Generally, the voice quality at the receiving end of a typical PCM system carrying voice traffic begins to deteriorate appreciably when the BER exceeds 10^{-3}, a point at which "cracking and popping" sounds may be heard due to bit errors. A BER of 10^{-6} is an acceptable minimum standard for a PCM system carrying 2400-baud data

channels. PCM systems are therefore designed so that under normal conditions, the BER lies well below 10^{-9} most of the time and only reaches a 10^{-6} threshold level for very small percentages of the time. The measured value of BER versus S/N ratio is shown in Figure 1. From the graph, it can be seen that a S/N ratio of 20 dB would correspond to a BER of 10^{-6}.

PCM vs. FDM

In a fade-free environment, typical of short paths or paths that traverse rough, nonreflective terrain in dry or elevated climates, low (near threshold) rf received signal levels may be specified for a PCM-over-microwave link. This is contrary to the engineering requirements of an FDM-over-microwave link, whose signal levels must be quite high even in such a suitable environment to provide the thermal noise quieting required to meet the high quality, low noise specifications of a modern communications system. As in cable systems, however, the BER

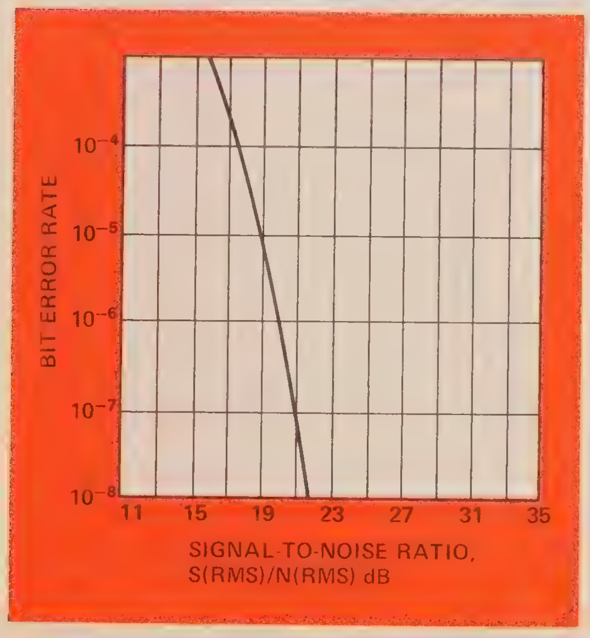


Figure 1. Measured values of BER versus S/N in the receive interface unit of the digital multiplexer.

in a PCM-over-radio system is affected primarily by the introduction of noise spikes or interference pulses into the data stream, which are decoded as legitimate bits of information (bit errors).

In a PCM-over-microwave system, error rates exceeding the 10^{-6} threshold criterion are introduced with a decrease in the signal-to-noise (S/N) ratio either by rf received signal level fading to threshold, or by high level co-channel or in-band rf interference. The S/N ratio related to a 10^{-6} BER assigned threshold, or outage value, is a function of the type of PCM or PCM-to-analog modulation coding (three level duobinary, PSK, QPSK, multilevel PSK, etc.) and the receiver detection techniques (discriminator, coherent detection, etc.) employed. The modulation coding and detection techniques employed in the GTE Lenkurt 2-GHz, PCM-over-microwave system permit the use of a standard analog FM radio system (the 78F2). In a typical BER-vs-S/N characteristic curve such as shown in Figure 1, the S/N ratio is directly related, dB-for-dB, to the rf received signal level and, in PCM radio systems, is of significance only near the radio noise threshold. An unusable BER point is reached only near the receiver threshold, and rf fades to threshold result from changes in the propagation medium (the atmosphere and terrain below it).

Microwave Propagation

The short wavelength of microwaves gives them many of the same properties of light waves; they are therefore refracted or bent by the atmosphere, and are obstructed or reflected by such obstacles as mountains, buildings, bodies of water, and atmospheric layers. While microwaves

travel at the speed of light in a vacuum, their speed in air is reduced and varies according to the varying density and moisture content of the air.

Gradual changes in air density may cause the radio wave to refract or bend continuously, so that the beam gradually curves toward the denser atmosphere. Because the atmosphere is increasingly thinner with higher altitudes, radio and, to a lesser degree, light waves, do not follow a straight path but are normally refracted downward. Radio paths extend beyond the visual "line-of-sight" horizon, since radio waves are more significantly affected by all three atmospheric density gradients (pressure, temperature, and humidity) than are light beams. The atmosphere is seldom homogeneous, but may be stratified or constantly changing (as evidenced by the twinkling of stars through what otherwise appears to be a distortion-free atmosphere). These atmospheric irregularities present a varying, nonhomogenous propagation medium to the microwave wavefront, which results in the propagation of not only the main body of energy but also many refracted and reflected secondary rays that arrive at the receive antenna at various phases and amplitudes. The amplitude of the resultant rf received signal—the sum of all of these main and secondary rays—could vary with time from 6 dB above normal to 40 dB or more below normal with signal cancellation. If the fade depth is greater than the fade margin provided in the transmission engineering process, a BER of greater than 10^{-6} is introduced into the PCM channel, resulting in an outage.

Fortunately, most fades of this magnitude result from atmospheric multipath and specular ground reflections, and are of extremely short duration. Only a small percentage of 2-GHz exchange plant PCM microwave paths would ever experience fading severe enough to require adding diversity protection or other special measures. Besides these short-term outages, long-term attenuation fades resulting from partial path obstruction or signal trapping may occur, but usually only in low clearance paths traversing shallow standing bodies of water (swamps, irrigated fields, lakes, etc.). An experienced transmission engineer can usually identify these unusual geographic conditions and route the microwave path over or around the suspected area.

The reliability (or availability) of a 2-GHz microwave path may be approximated with a suitable degree of accuracy from weighted Rayleigh distribution curves as shown in Figures 2A and 2B. These curves show that short 2-GHz microwave links only rarely experience a fade of such depth as to cause an outage, whereas longer paths may be subject to deeper and more frequent fading. The locality is of considerable importance, as shown in Figure 2B; long 2-GHz paths traversing reflective terrain or water in coastal or other highly humid regions will fade far more frequently than long paths over rough terrain in dry regions.

Most applications of the short haul PCM-over-microwave links are configured for single-channel operation, with no need for equipment or propagation redundancy. In longer systems in difficult propagation areas, space diversity may be used. Figure 2B shows that the outage time for a 30-mile path with 35 dB fade margin may be reduced from 48 to less than 2 minutes per year with a 40-foot diversity spacing of the receive antennas.

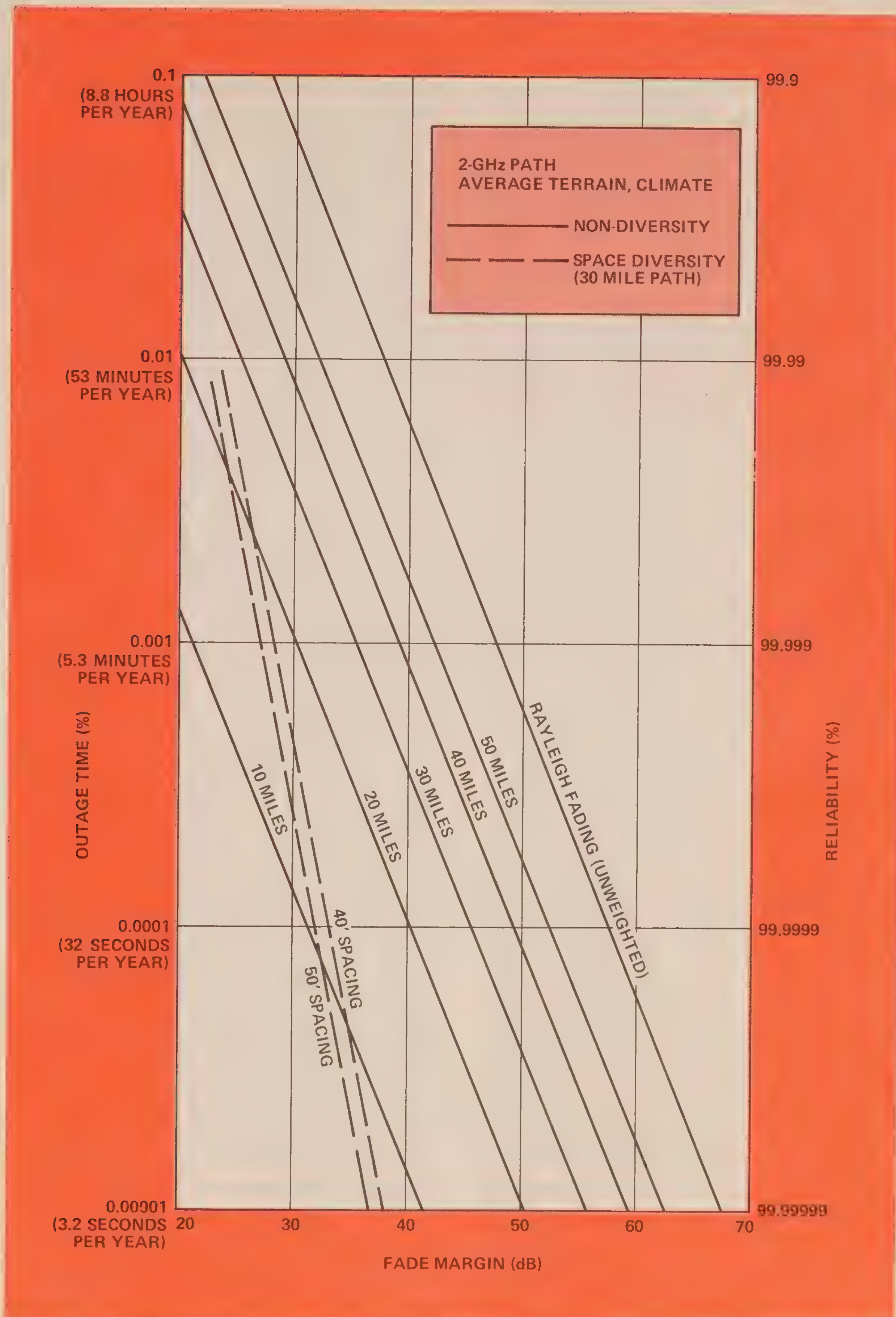


Figure 2A. Outage time due to fading, as a function of path length, can be estimated using weighted Rayleigh distribution curves.

Receiver Noise

In the receiver, the level fluctuations of the received rf signal caused by fading are removed by automatic gain control (agc) circuits before the signal is applied to the demodulator. In most microwave receivers, agc is provided at the intermediate frequency (70 MHz) to which the received signal is converted by the mixer. Thus, the receiver gain varies in accordance with the received signal level, the gain being high when the received signal is faded and low when it is not. Any noise entering the receiver input, as well as noise generated in the input circuit components, is amplified along with the desired signal, so that when the signal fades, the noise is proportionally higher and the signal-to-noise ratio is decreased.

In terrestrial line-of-sight microwave transmission, the limiting noise contributor is the thermal background noise of the warm earth (typically −114 dBm per MHz of receiver bandwidth). Added to this background noise is the receiver front end noise, characterized by the receiver noise figure indicating (in dB) how much more noise is applied to the receiver mixer compared to the −114 dBm/MHz warm earth background. Typical receiver noise figures range from 4 to 12 dB depending on receiver front end design and frequency band. The PCM signal must be 10-20 dB above the thermal noise level (a value determined by PCM coding and detection techniques) for a given threshold error rate of 10^{-6}.

Digital Transmission

A discussion of microwave propagation applies to any type of point-to-point microwave system, whether carrying digital or FDM channels. The information to be transmitted may be carried by the microwave signal in various forms of modulation such as frequency, phase, amplitude, or a combination of these; but, frequency or phase modulation is the preferred method because it facilitates provision of agc, and amplitude linearity is not required in the rf and if circuits.

When the information to be transmitted is in digital form—digital data or PCM voice, for example—the preferred method of modulation of the microwave carrier is still frequency or phase modulation, thereby keeping the envelope of the microwave signal as constant as possible. In this light, it can be seen that FM microwave equipment originally designed to transmit such analog signals as FDM voice or video, should also be suitable for the transmission of digital information. Everything necessary for satisfactory and reliable transmission is already present. All that is required is conversion of the digital signal into a form in which it can take the place of the normal analog modulating signal; that is, make it look similar with respect to level and frequency spectrum.

A typical example of this approach is GTE Lenkurt's 9120A digital multiplexer, which contains facilities to interface two low speed T1 PCM bit streams (24 vf channels each) with an existing FM radio of suitable bandwidth (such as the GTE Lenkurt Type 778F2A microwave radio operating in the 2-GHz common carrier band). A typical arrangement of a 2-GHz, PCM-over-microwave system is shown in Figure 3. At the transmitting end, the 9120A combines two asynchronous DS1 (1.544 Mb/s) bit streams into a single output unipolar digital signal of 3.156 Mb/s. This signal is converted into a "modified duo-

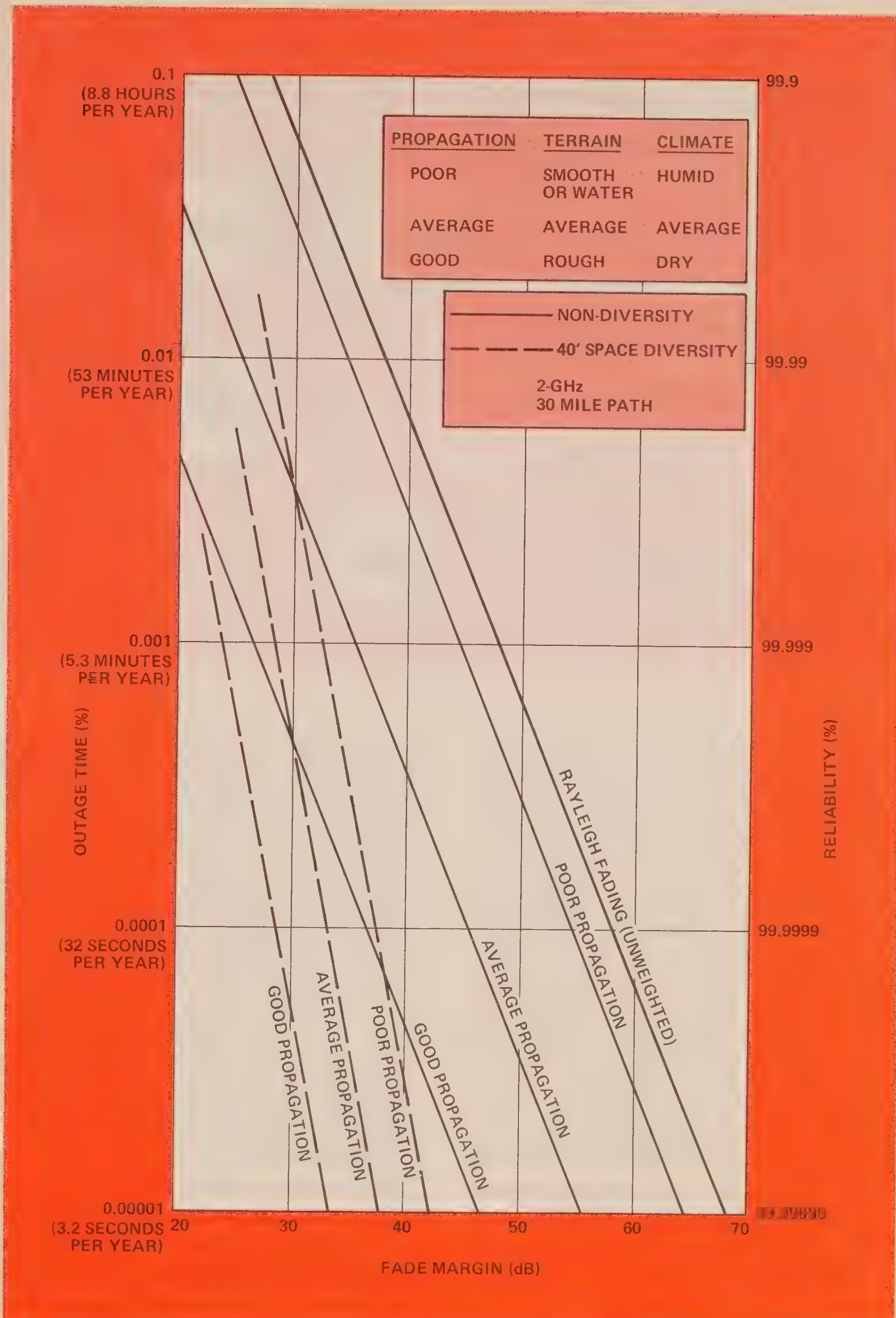


Figure 2B. Climate and terrain are important factors in determining overall microwave system performance.

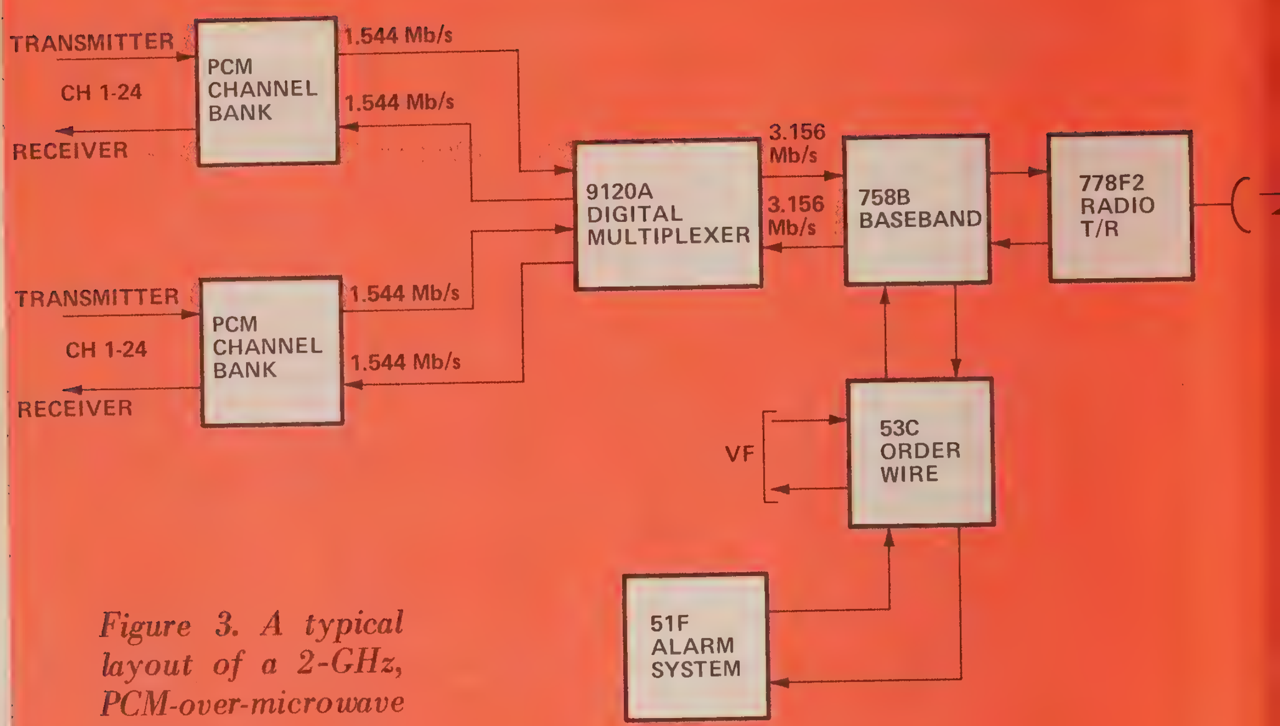


Figure 3. A typical layout of a 2-GHz, PCM-over-microwave system.

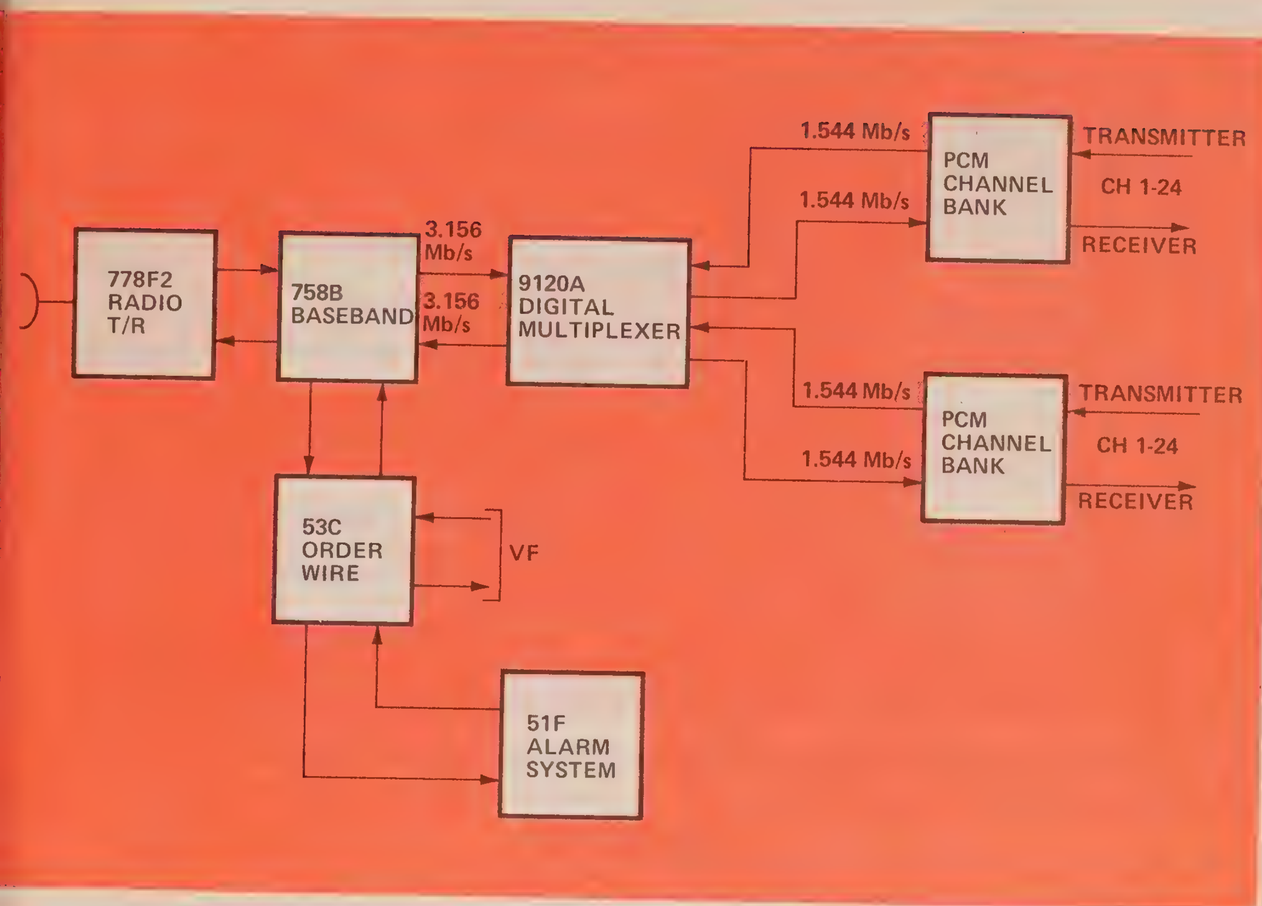
binary" signal in the radio-interface unit and applied to the baseband input of the radio, where it frequency modulates the microwave carrier.

At the receiver, the modulating signal is recovered in the conventional manner by the FM discriminator and then applied to the receive radio interface unit of the multiplexer. The interface unit extracts the clock from the modified duobinary signal and subsequently converts it back into a digital (binary) signal for further processing (de-multiplexing). Finally, the two original 1.544 Mb/s bit streams emerge from the multiplexer for connection to such T1-type equipment as channel banks, data terminals, or repeatered lines.

Fade Margin

Fade margin is one of the important factors that determine microwave system performance. Fade margin is the amount of reserve power in dB available at a receiver to overcome the effects of sudden atmospheric fades. If, for example, the normal receiver input level is –45 dBm, a 30-dB multipath fade will drop the receive level to –75 dBm. If the –75 dBm level corresponds to a S/N ratio in the receive interface unit of 20 dB and a BER of 10^{-6}, then any further increase in the intensity of the fade will exceed the limits of the bit error rate. For the conditions just described, the fade margin is 30 dB.

The fade margin depends on the arrangement of the system. The closer the transmitter is to the receiver, or the more suitable the climate and terrain is for microwave propagation, the smaller the fade margin necessary for required reliability. The minimum rf-transmit output power of the GTE



Lenkurt 778F2A radio transmitter (without hot-standby) is +36 dBm. The net path loss of a system is the total loss inflicted on the signal throughout the microwave path. It includes the loss of the path, the gain of the antennas, the loss in the coax between the antenna and the radio, and anything else in the rf path between the transmitter and receiver antenna ports.

A typical 2-GHz net path loss might be 71 dB if the transmitter and the receiver are approximately 30 miles apart, and 8-foot parabolic antennas are used. Such a loss means that normal receive carrier power will be approximately –35 dBm (+36 dBm – 71 dB = –35 dBm). The fade margin in this example is 40 dB, which means that the signal can drop 40 dB without exceeding 20 dB S/N ratio and a 10^{-6} BER.

The BER of a PCM microwave system is comprised of the total contribution of the digital multiplexer, repeatered line (if used), and the radio. The radio's threshold error contribution varies with the carrier-to-noise (C/N) ratio as well as the signal-to-interference (S/I) ratio at the microwave receiver. The C/N ratio varies with fading, and a 1 dB increase in noise will typically result in a ten-fold increase in errors. A safety or fade margin is required to avoid exceeding the error rate specified in the system reliability objectives more than a certain percentage of the time. The fading characteristics of a particular microwave path, the required propagation reliability, and the use or omission of diversity protection, determine the necessary fade margin.

The detected signal level is held constant by agc and amplitude limiting

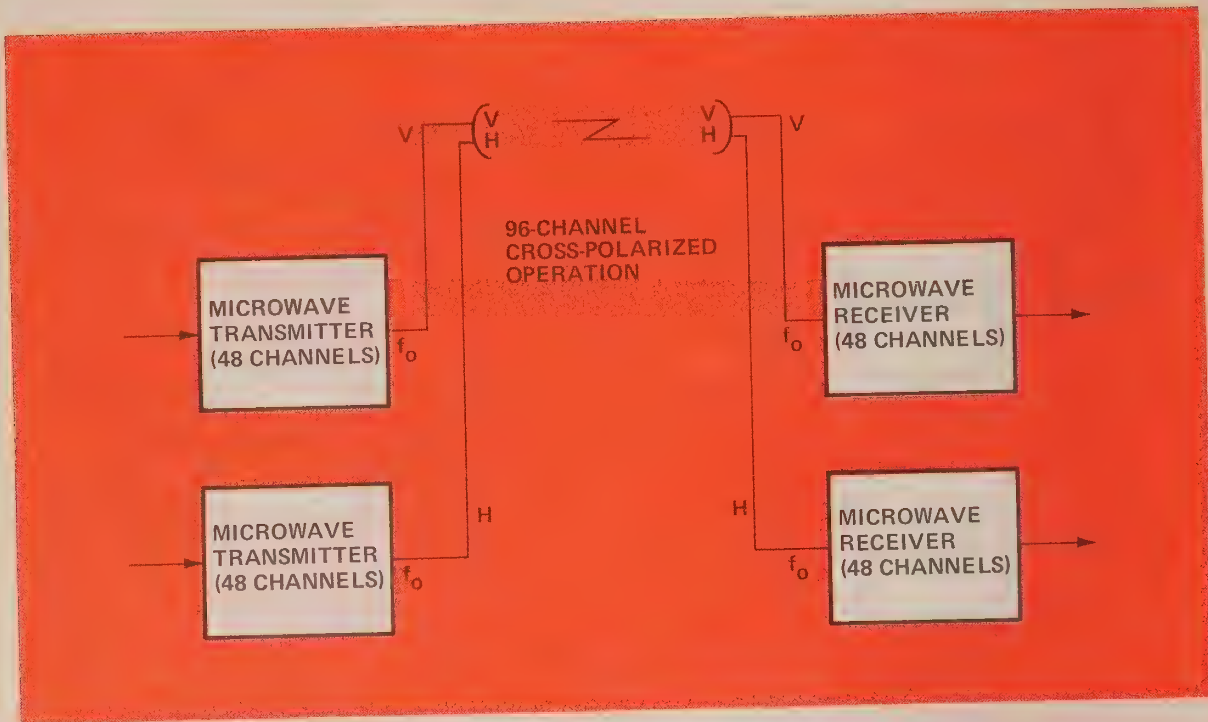


Figure 4. The use of cross-polarization can double the channel capacity of a digital transmission system (Figure shows one direction of transmission only.)

in the if section of the radio, and the noise level may vary over a wide range as a result of microwave signal fading. Over the normal operating range of rf input levels, the error rate is very low. In contrast to a conventional FDM system, noise under unfaded conditions or during moderate fades does not significantly degrade PCM system performance. The digital system is far more tolerant of interference, with error rate being significantly affected only when the power of the interfering signal approaches that of the noise at the error threshold. And, unlike frequency-modulated radio systems carrying FDM channels, the effect of interference on the digital signal is largely independent of the spectral characteristics of the interfering signal.

Cross-Polarization

Because PCM transmission over microwave radio is less sensitive to interference than is FDM transmission, both horizontal and vertical polarization of the same radio frequency can be used for two independent PCM systems over the same path in most propagation conditions, thereby doubling the route capacity. Typically, the value of XPD (cross-polarized discrimination) of a carefully aligned antenna system provides a 25- to 30-dB polarization advantage on a single hop. This means that the vertically polarized signal is attenuated 25 to 30 dB from the horizontally polarized signal (or vice versa). While this separation is unacceptable for transmission of FDM signals, it has an insignificant effect on the transmission of digital signals except near the receiver threshold. Both the horizontal and vertical polarization transmission of a single radio frequency can therefore be used for digital transmission. Thus, two signals from different radios occupying the same rf frequency assignment, but representing two different digital

systems, can be transmitted over the same antenna and the same path (see Figure 4). Through the use of cross-polarization, the digital transmission channel capacity over 2-GHz radio using the 9120A digital multiplexer can be increased to 96 channels per radio frequency by using two radios.

While the use of multifrequency cross polarization has been an effective means of transmission for many years, it is not without limitations in single-frequency applications. The XPD may diminish during multipath fading, decreasing the S/I ratio to considerably less than the 20 dB that equates to a threshold of 10^{-6} BER. Similar loss of XPD has also been suspected to result from heavy rainfall, which otherwise is not a factor in 2-GHz propagation characteristics.

Frequency modulation (FM) techniques have been consistently used in commercial microwave radio systems for multichannel voice and data transmission. These systems are used in conjunction with the familiar frequency division multiplex (FDM) carrier system. Proven FM radio equipment can now offer an economically attractive alternative for the extension of digital systems over obstructions or into remote areas where cable plant costs are prohibitive. This three-part series has presented some of the many aspects of a new technique now available to meet the ever-expanding needs of the telecommunications industry.

BIBLIOGRAPHY

1. *Digital Transmission Over 2-GHz Radio System*, San Carlos, California: GTE Lenkurt Inc., 1973.

2. "Microwave Diversity. . .how it improves reliability," *The Lenkurt Demodulator*, March, 1961.

3. Olson, J.D., "A Method of PCM Transmission Over 2-GHz Microwave Radio," *International Conference on Communications*, (June 1974), 33C-1 – 33C-5.

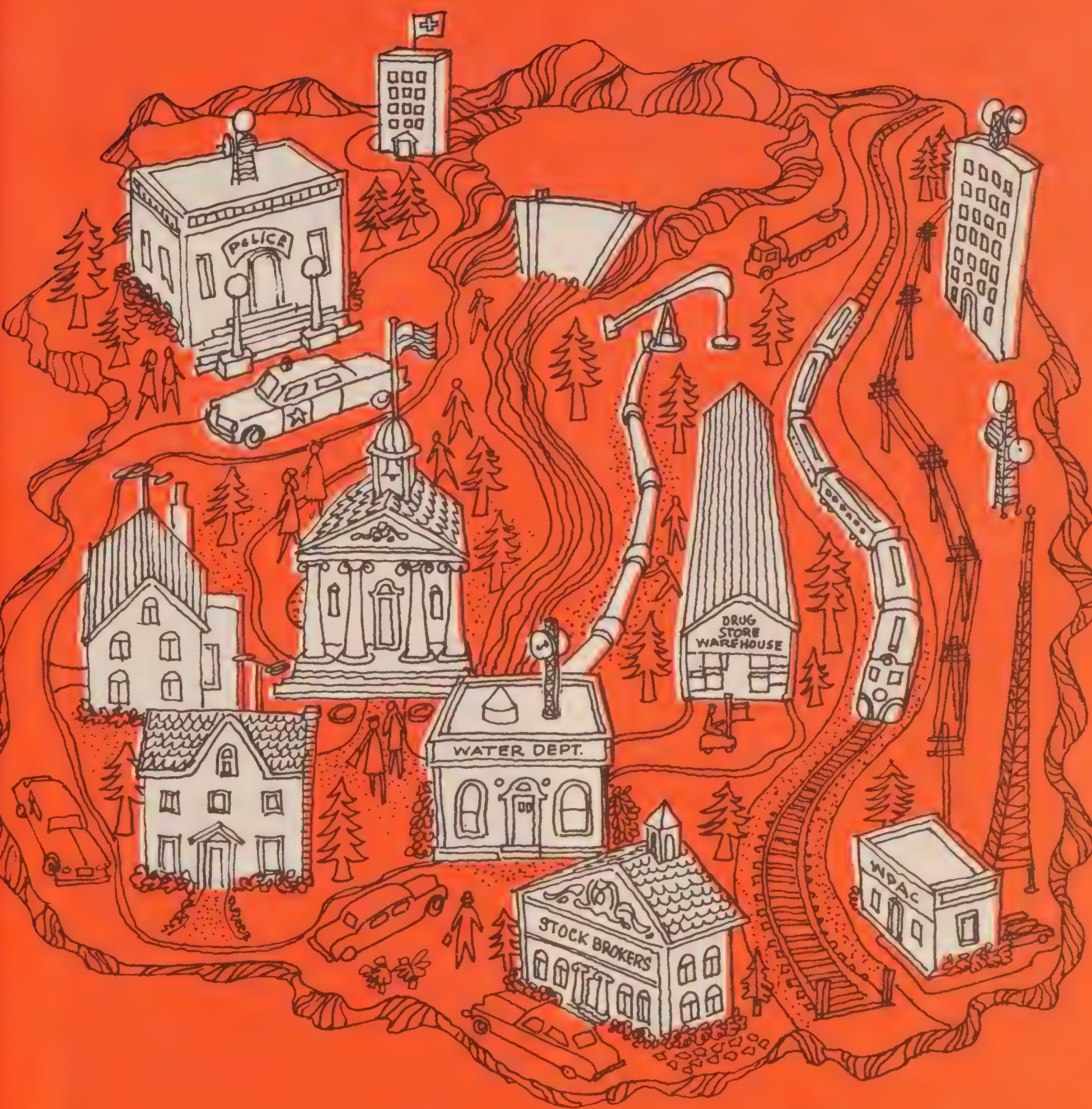
4. Swartz, L.T. "Performance Analysis of a Three-Level Modified Duobinary Digital FM Microwave Radio System," *International Conference on Communications*, (June 1974), 5D-1 – 5D-4.

5. White, Robert F. *Engineering Considerations for Microwave Communications Systems.* 2nd Ed. San Carlos, California: GTE Lenkurt Inc., 1972.

SECTION III
MICROWAVE RADIO

GTE LENKURT DEMODULATOR

MAY 1972



microwave applications

Each microwave communications network is designed to fit its specific application.

Microwave refers to different things in different industries. To the telecommunications industry it refers to radio relay systems operating at microwave frequencies (above 900 MHz). Within the telecommunications industry, it can mean communications over distances as short as a mile for urban applications, or as long as 45,000 miles for communications via satellite. Microwave systems can be used to communicate over bodies of water, over mountain peaks, across deserts, or across the street. And microwave can be used to transmit many forms of information – spoken words, written words, music, photographs, data pulse streams, or television programs.

Microwave systems have the advantages of large information-handling capability, high reliability, low power consumption, good protection against weather, and low maintenance costs. Disadvantages include short transmission range (from one earth antenna to another) and susceptibility to interference by objects obstructing the transmission path.

The information handling capability of a microwave system – the number of channels, data transmission rate, usable bandwidth, and similar aspects – is dependent on the performance quality required, and the stability of the transmission equipment. Performance quality is the degree of freedom from the noise and distortion which obscure signals or tend to create transmission errors.

Reliability refers primarily to freedom from equipment failure, since protection against signal fading is largely a function of system or path engineering and the use of such techniques as space diversity.

Even if the question is whether to lease channels from a common carrier or to build a new transmission system, the answer can only be found through consideration of a wide range of specifics, some common to all systems, some unique to one system alone. Systems are tailored to the required information handling capacity and the type of information to be transmitted, as well as the physical surroundings or environment in which the system is

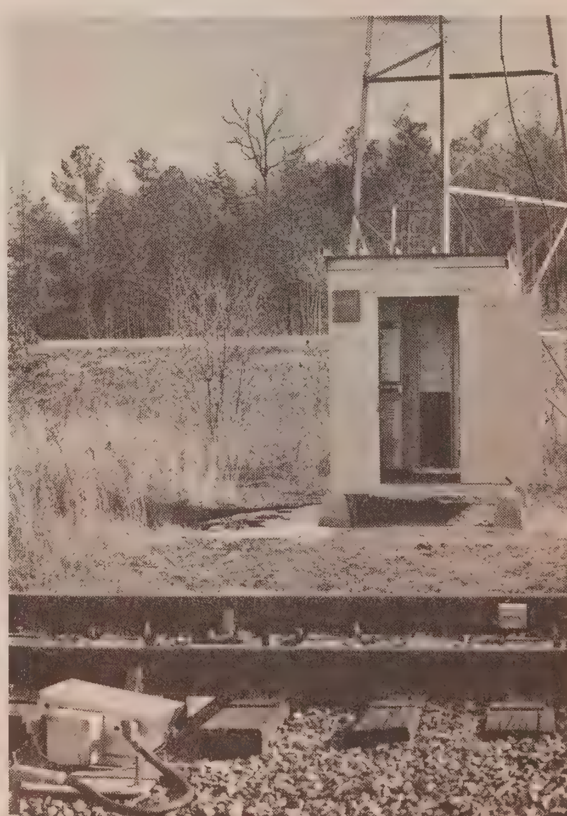


Figure 1. Microwave networks are an integral part of the railroads' hot-box detection systems.

Figure 2. Microwave repeaters operate unattended from the tops of mountains.

expected to operate. For this reason, each system is unique in its application yet universal in its operation.

Increased Industrial Use

Microwave communications were first used commercially in 1947 to transmit voices between Boston and New York. Until recently, telephone companies were the most common commercial users of microwave for long-distance video, voice, and data transmission networks. But, more and more industries are beginning to use microwave transmission for their own communications.

A tremendous spurt to industrial usage occurred in September 1960 when the FCC (Federal Communications Commission) made microwave frequencies available to virtually any type of business or industry in the United States.

Railroads are typical among the long-time industrial users of private microwave systems. Microwave networks are used to insure the safe, smooth operation of the railroads. The status of shipments and the locations of locomotives, freight cars, and piggyback trailers is available instantly through the use of a microwave network tied to a central data processing center. An average day on the railroad might result in as many as 20,000 messages concerning moving railroad cars.

Hot-box detection systems are used to locate hot journal bearings on moving trains. These systems (see Figure 1) transmit information on the location of any hot journals to a central processing center. Mobile telephones are then used to notify train crews along the main line in order to stop trains with overheated journals. This same network can also be used to detect equipment dragging from freight cars, and to notify the train crews so that the situation can be corrected.

Microwave communications are being used for growing ACI (automatic car identification) networks that aid the railroads in computerized traffic control. Accounting for all rolling stock on one railroad is a monumental task. The task is compounded by the number of railroad companies in any one area and the free access to each others cars for handling freight. With ACI, cars are counted and identified as to owner, type and serial number. This information is automatically forwarded to a data processing center while the trains keep moving at speeds

Figure 3. Microwave systems are used for inter- and intra-island communications in Hawaii.

of up to 80 miles per hour. This system can also be used for recording cargo, destination, and direction of travel.

Since hot-box data transmission, ACI information, teletype and data transmission, and reliable voice communications are a vital part of smooth railroad operation, microwave networks are ideal for these applications.

Remote Installations

Long-haul microwave is dependable under all conditions, even when operating unattended in remote locations (see Figure 2). The reliability requirements of most microwave systems are exceedingly stringent – as close to 100% as economically possible. Systems can be provided today with propagation reliability in excess of 99.99% per hop. This kind of reliability can be expensive, but in some systems, such as pipeline control networks, the cost can be justified to get higher reliability.

Complex control networks are used in the pipeline industry to control both production and petroleum flow. Drilling and pumping operations can be remotely controlled through the use of microwave networks. Remote chemical analysis of production samples is used to control drilling. When the percentage of petroleum in a sample drops to a predetermined value, production is stopped at that well head, permanently or until further drilling locates a new source of fuel.

Along the pipeline itself status monitoring and remote supervisory control are vital to the success of the total operation. Volumetric output, flow rate, pressure, and temperature are monitored constantly. Hundreds of miles from the nearest civilization, valves are opened and closed, and pumps are automatically activated or stopped.

Although techniques and equipment vary considerably from one pipeline complex to the next, the conditions monitored and the functions performed are essentially identical – regardless of whether it is natural gas or crude oil being gathered and transported. Petroleum pipeline networks crisscross some of the most remote and rugged terrain in the world.

High, rugged mountains can actually help, rather than hinder microwave transmission, since mountain peaks permit repeater sites to be spaced further apart. Thus tower heights can

Figure 4. Ice and snow do not restrict the usefulness of this microwave defense network.

be kept to a minimum. On flat land, tower heights have to be increased to allow for earth curvature, and so that antennas can rise above man-made obstructions.

Large bodies of water do not hamper microwave transmission, either. Microwave systems are used extensively throughout the Hawaiian Islands for voice communications (see Figure 3). In fact, in some cases, communications are established by using microwave links back and forth between islands for the ultimate purpose of intra-island communications. The wire lines that had been previously used for communications within the island proved unsatisfactory since they were difficult to maintain and were sensitive to the tropical weather.

Microwave equipment is made to withstand severe weather conditions. Antennas, for example, can be heated to protect them from damage caused by ice and snow (see Figure 4). Bad weather areas, such as those subject to flooding, are particularly appropriate for microwave systems — evidence of this is offered nearly every year when microwave takes over in times of crisis to provide the only means of communicating with stricken areas.

Large bodies of water are also controlled through the use of microwave. Electric utility companies use microwave systems to remotely control the water height behind dams at hydroelectric generating stations (see Figure 5). Microwave networks are also used along transmission lines (see Figure 6) to signal a control center if the current flow becomes abnormal for reasons such as higher than normal demand, downed wires, or generator malfunction.

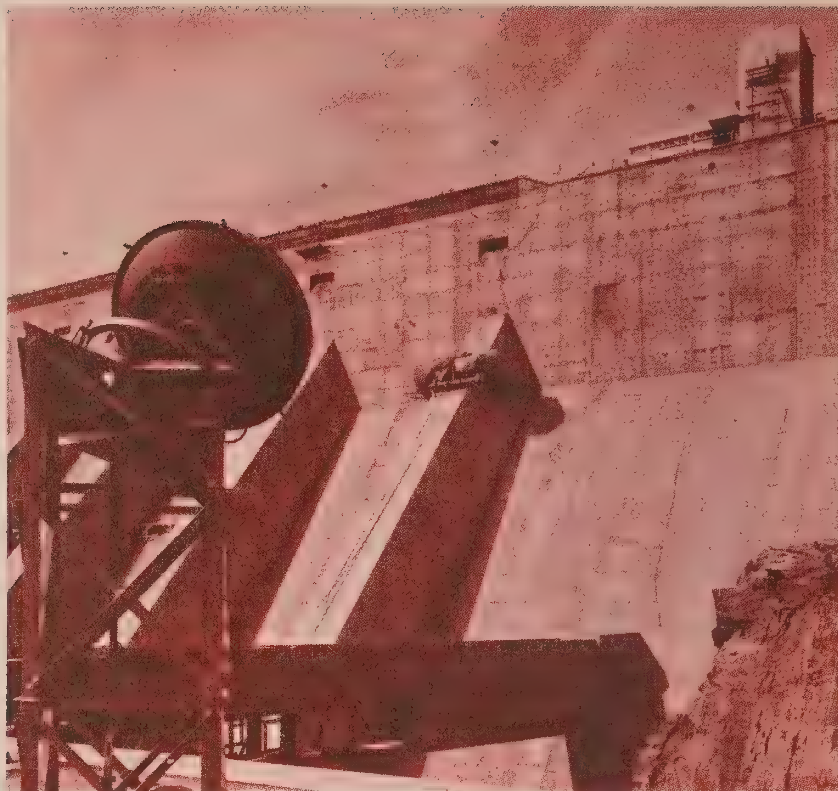
Added Advantages

Microwave communications offer certain distinct advantages over other transmission media, especially for industry or business installing its own private system. Unlike other communications systems, microwave does not require right-of-way along the total transmission path. At the same time, the installation of a microwave system still requires the acquisition of land — for terminals and repeater sites — and these areas must be provided with access roads.

The electric utilities, pipelines, and railroads own the right-of-way along their system networks. But business and private industry with scattered operations, as a rule, do not own a path of land from one office to another. Microwave, with its limited need for land acquisition, is the most promising form of private communications for business and industry.

Microwave transmission also offers a high degree of privacy — eavesdropping would have to be done right on the signal beam using electronically identical equipment. An added feature is that other systems can use the same frequency so long as they don't use the same transmission path.

Figure 5. Microwave systems are used to monitor and control generating operations at hydroelectric dams.



It is believed by many that the computer and its remote data terminals will influence human life for many decades to come. The privacy offered by microwave transmission, as well as its large information handling capacity, make microwave ideal for linking these data terminals together.

Perhaps one of the most common data networks is used by engineering groups in industry to access time sharing computer services for design work. New data networks are being established daily. Commercial banks need extensive data communications networks to keep up with the growth of credit card authorization, and for their traditional data exchanges regarding money flow. Stockbrokers and security exchanges also have to constantly keep track of money flow throughout the country and the world. Microwave systems can provide private intra- and inter-city, as well as inter-country data service for these new users.

Manufacturing plants with geographically scattered operations are finding private microwave networks useful for such things as process control, production control, inventory control, and invoicing. Many locations can be tied together by a central data processing operation.

Retail chain stores can also use private microwave networks connected to a central warehouse to make sure their shelves are properly stocked, while minimizing the need for extensive warehousing at each store. Such special purpose communications networks lead to more efficient operations and lower overhead.

Information Networks

Extensive microwave networks are used by law enforcement agencies for rapid information retrieval from centrally located files. These networks also tie together mobile base stations that are used to keep in contact with law enforcement vehicles. Instant communications are essential in law enforcement, and regional and nationwide microwave networks are being established for this purpose.

Figure 6. Electric utilities use microwave systems to monitor current flow along transmission lines.

Another form of information transmission is needed for educational purposes. Microwave installations are used to transmit university classes to industry for cooperative educational arrangements. These remote classrooms have two-way voice communications as well as video communications so that the students at both ends can participate in classroom discussions. Medical schools also use microwave installations for video transmission. Such installations permit medical students to "participate" in surgical operations without leaving their classrooms and also keep practicing doctors up to date on the latest developments in their field.

The general public can be kept informed with live radio and television news coverage of all kinds. Portable microwave systems provide wide news distribution. These systems operate from small vehicles that can be taken right to the scene as easily as taking a film crew for the same news coverage.

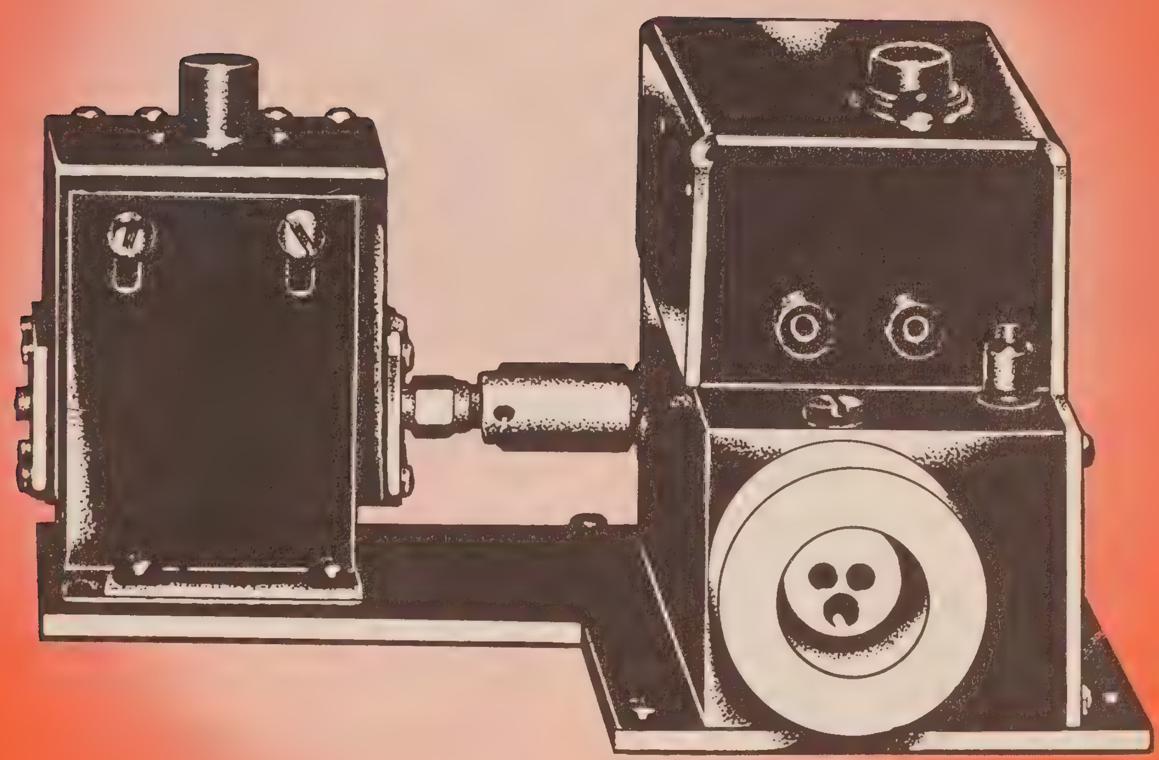
Microwave communications have come a long way since 1947. Today they enable people to communicate at incredible speeds anytime, anyplace, in almost any form – still and motion pictures, sophisticated computer language, heartbeats of astronauts in outer space. It speeds news copy and photos to newpapers across the country; broadcasts television programs to hundreds of cities; links military posts throughout the world; rushes diagnostic information from patients to doctors; reports the presence of unidentified aircraft to defense centers; connects bedridden students with their teachers and classrooms; and even raises and lowers drawbridges.

System operators who have sampled the benefits of microwave are adding more links to their facilities as expansion is required. And, new uses are continually being found for microwave. It is sufficient to say that microwave is widely recognized as a flexible, reliable, and economical means of providing point-to-point communications.

GTE LENKURT

DEMODULATOR

JANUARY 1973



innovations in microwave transmission

Recent developments in solid state devices and transmission control have had a profound effect on microwave radio equipment, with especially pronounced results in baseband radio.

In the telecommunications industry, microwave radio systems are used to transmit wideband, baseband signals. Such baseband signals, which may consist of multichannel telephony, video, or wideband data, are transmitted between microwave terminals. Where the distance between terminals is such that line-of-sight transmission is not feasible, microwave repeaters are used at intermediate points between the terminals.

Generally, the types of radio equipment used both at microwave terminals and microwave repeaters can be divided into two categories – baseband (directly modulated) and IF (intermediate frequency) heterodyne. Each provides a different way of modulating the microwave carrier to convey information from one location to another, and each is designed for a specific type of application. IF heterodyne systems are usually used in long-haul networks where the information conveyed does not need to be dropped off at repeater points along the way. Baseband systems are generally utilized in short- and medium-haul applications, where access to the baseband is necessary along the microwave route. In the past, however, an IF system has sometimes been chosen – because of its better noise performance – to do a job which might otherwise be accomplished by a less expensive baseband system. Today, with modern technological improvements, baseband systems have been enhanced in performance to closely approximate IF heterodyne performance, in short- and medium-haul routes.

In IF as well as baseband systems, the receiving equipment performs essentially the same function; the great difference between the two systems is manifested in the nature of the transmitting equipment. And, it is in the baseband transmission equipment that significant changes have occurred. One major change has come as a result of the incorporation of solid state microwave sources into baseband radio equipment; but other changes, such as improvements in transmission control,

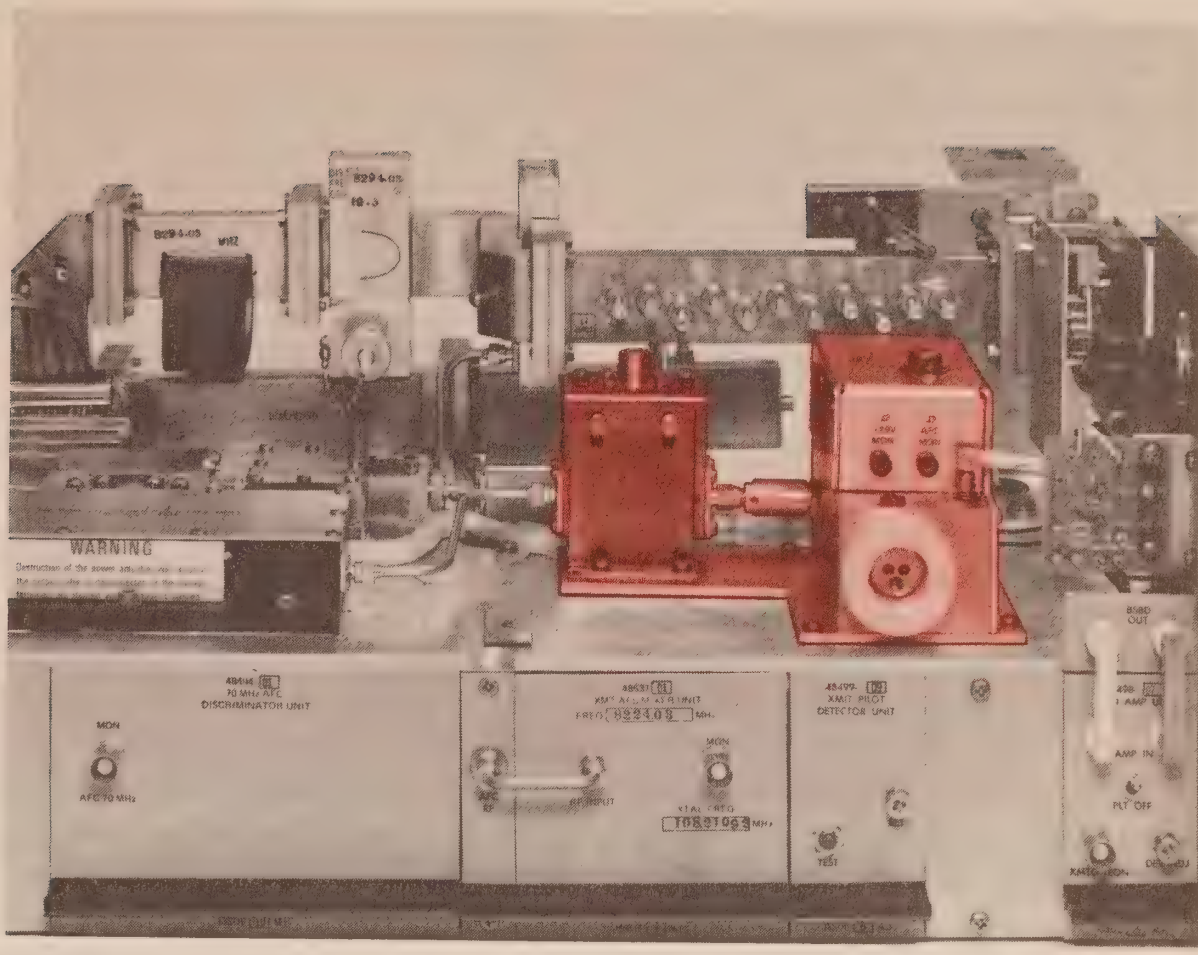


Figure 1. The replacement of the klystron with a solid state microwave source is one of the major changes in GTE Lenkurt microwave radios.

delay networks, and linearity, and the use of more stable components throughout, have been instrumental in improving the performance and reliability of baseband equipment. The advances in solid state technology have been beneficially applied to both baseband and IF systems, but the most dramatic changes have occurred in baseband radio, where there was the greatest margin for improvement.

Microwave Sources

The microwave source is the device which determines the frequency of transmission and reception. (See Figure 1.) In baseband transmitters, where it is directly modulated by the baseband signals it is known as the FMO (frequency modulated oscillator). In heterodyne transmitters, where the microwave source output is mixed (heterodyned) with the baseband modulated output of the MODEM (modulator-demodulator), it is generally known as the *up converter.*

In all commonly used microwave receivers (for either IF or baseband systems), the microwave-source output is mixed with the received microwave signals to produce an intermediate frequency. Here the microwave source

is also known as the *down converter, or local oscillator*. In this case it determines the frequency of reception by having the IF frequency (usually 70 MHz) either added to, or subtracted from its own frequency. Previous to the introduction of solid state microwave sources, the klystron tube, which required extremely high voltage power supplies, was used as the frequency generator for both baseband and IF heterodyne systems.

Historically, baseband microwave transmitters were developed to take advantage of the reflex klystron (developed for use in radar equipment during World War II), the characteristics of which made it ideally suitable for direct modulation, since the modulating signal (baseband) could be directly applied to the klystron reflector. This process produced a relatively uncomplicated transmitter in that the klystron, with its frequency-controlling cavity, was, in fact, the entire transmitter. The disadvantages of this arrangement were primarily the problems of frequency stability, the necessity of providing the stable, high-voltage power supplies required to operate the klystron, and the problems associated with producing modulating amplifiers with a high degree of linearity while, at the same time, delivering a relatively high modulating voltage output.

The frequency stability problem was, to a degree, solved by using stable reference cavities with automatic frequency control (AFC) to the klystron. Later, the inherent stability of the klystron itself was further improved by stabilizing the temperature of its frequency-controlling cavity, the physical dimensions of which determined the operating frequency. Still, the linearity requirements for the modulating amplifiers, and the high voltage power-supply requirements remained a source of concern to telecommunications equipment manufacturers.

Further Advances

The next step in the development of microwave transmitters resulted from the development of efficient solid state, frequency-multiplying devices. With the advent of these, it became possible to have the baseband signals phase modulate, amplify, and multiply the output from stable, crystal-controlled oscillators operating in the 100-MHz region. The problems associated with this arrangement were those of (a) producing a relatively high-power output at the crystal-controlled frequency to offset the power loss associated with each frequency multiplication and (b) noise introduced into the modulated RF output by the many multiplication stages. The high-voltage power supply, the high-voltage swing of the modulating amplifier, and the frequency stability problems associated with the klystron transmitter were, however, largely solved. The noise performance of the transmitters remained essentially unchanged.

The more recent development of the AFC controlled, solid state microwave source has dramatically reduced the number of frequency multiplying stages required (from about 60:1 to 4:1 in the 6- to 8-GHz bands), thus

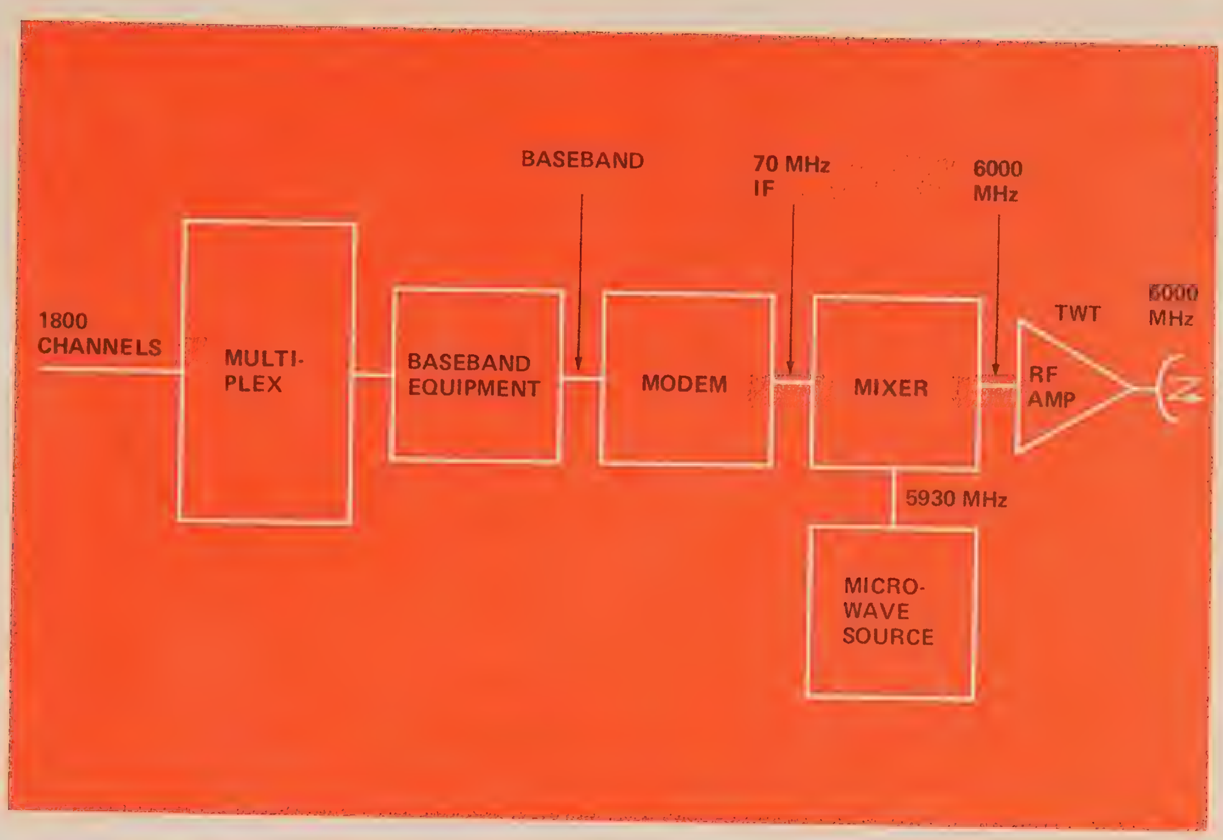


Figure 2. The transmit section of an IF heterodyne microwave radio terminal connected to telephone multiplex equipment.

reducing the total noise from these devices.

Amplification at Repeater Stations

One of the major problems in microwave transmission is that of amplification at microwave frequencies. The simplest non-remodulating repeater is the RF type in which the incoming modulated RF signal is only amplified and translated to the outgoing radio frequency. The difficulties associated with the stable amplification of microwave frequencies with existing devices, have, however, so far prevented extensive use of RF repeaters.

To better appreciate the problem of microwave amplification, it should be recalled that a typical receive-signal level on a microwave system is of the order of –30 dBm, and a transmitter output is typically around +30 dBm. Working systems must also take the effects of fading of the receive signal into consideration. Typically, fades of 40 dB may be expected so that received signals as low as –70 dBm must be considered.

The difference between these two power levels (–70 dBm and +30 dBm) represents the gain necessary at a repeater. This is 100 dB, or a power amplification of 10,000,000,000. While stable solid state devices needing these high gain requirements are readily available for amplifying frequencies of the order of 70 MHz, they are not

yet generally available for microwave frequencies. So, it becomes necessary to reduce the microwave signal frequency (from 6 GHz, for example) to around 70 MHz (the intermediate frequency), by mixing it with the output from a microwave source. At 70 MHz, the required amplification can be achieved with readily available and well proven devices which operate from low-voltage dc supplies. The output from this IF amplifier carries all the baseband information from the received microwave signal at a greatly increased power level but at a much lower frequency. It is still, however, a frequency modulated signal and it is the manner in which this signal is treated before onward transmission where the basic difference between IF and baseband repeaters lies. In the heterodyne repeater, the IF signal is mixed with the output from a second microwave source, and the resulting RF signal is usually amplified and fed to the transmitting antenna. In a baseband repeater, the modulated IF signal is demodulated and used to remodulate the microwave source (FMO) of a baseband transmitter.

IF Systems

Baseband and IF heterodyne microwave radio systems differ in the way they process the baseband; baseband meaning the entire volume of information to be transmitted by the microwave system. This might, for example, be 1800 voice channels coming from the multiplex equipment of a telephone company. Microwave terminals provide the facilities for modulating and demodulating the RF carrier to permit the insertion and retrieval of the baseband signals.

At an IF transmitter terminal (shown in Figure 2), the baseband coming from the multiplex equipment is first adjusted to the correct signal level by the baseband equipment. It then modulates (deviates) a 70-MHz oscillator contained in the modulator section of the modem to produce the frequency-modulated IF signal. The modulated intermediate frequency is then heterodyned (mixed) with the output from a fixed-frequency, solid state microwave source. The desired product of the mixing process (either the sum or difference of the 70 MHz and the microwave source frequency) is filtered out and usually amplified by a TWT (traveling wave tube) before being applied to the transmit-antenna system. The function of the microwave source in this process was once performed by a klystron tube.

At through IF repeaters (see Figure 3), the received RF signal coming from the terminal, is heterodyned with a microwave source and filtered to produce the IF frequency, which is amplified, heterodyned with a microwave source, and filtered to produce an output at the repeater transmit frequency. This is usually amplified again before being fed to the forward direction antenna. The 100-MHz difference between the input to the repeater and the output, is typical of the frequency necessary to avoid interference between the two. This also applies to baseband repeaters. From this description, it can be seen that microwave sources play a major role in microwave transmission systems. The important

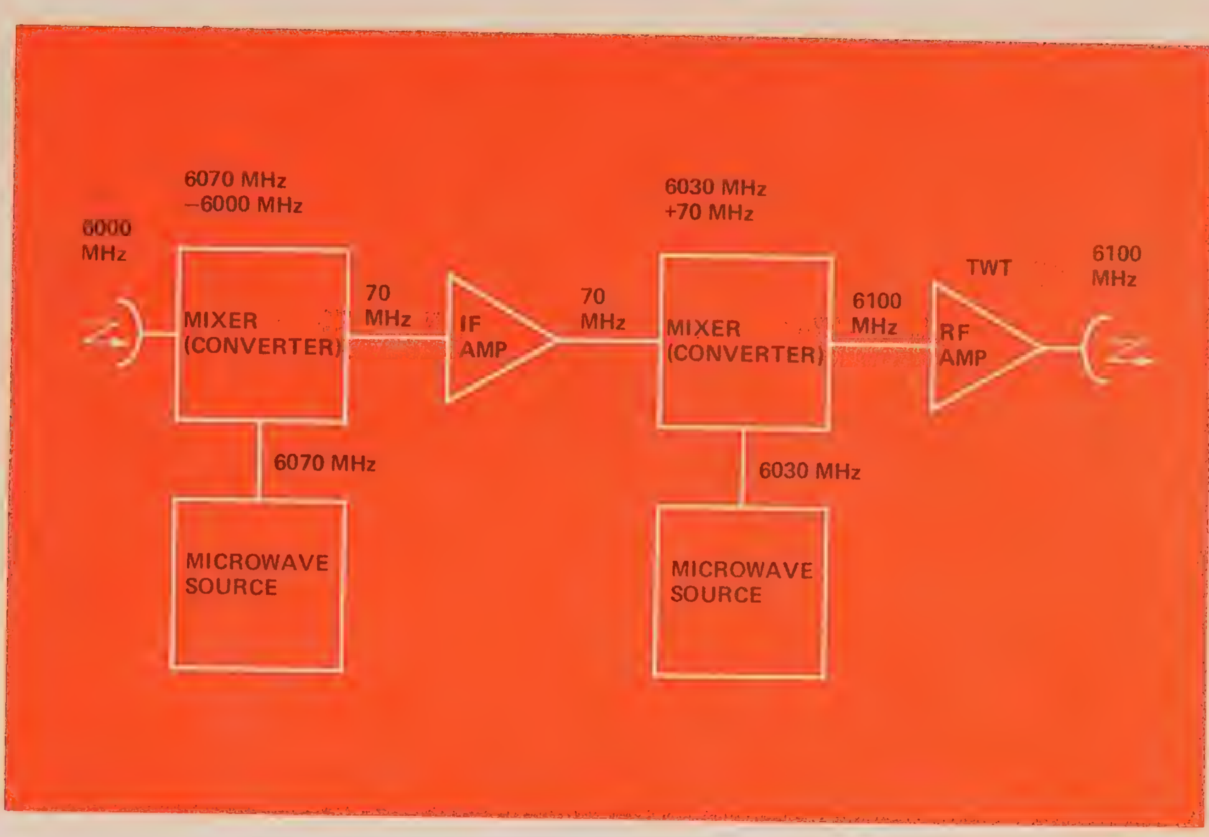


Figure 3. An IF repeater does not *require extraction of the baseband from the RF carrier during amplification.*

thing to remember about IF systems is that their great effectiveness lies in that, ideally, the RF carrier is not demodulated down to baseband at any point along the route. If it becomes necessary to access the baseband at any repeater, noise caused by the demodulation-remodulation process is introduced, and the system begins to lose its effectiveness.

Baseband Systems

The "baseband" microwave system is so termed because at each repeater stage, the incoming modulated carrier wave is demodulated down to the baseband frequency.

In a baseband terminal, such as GTE Lenkurt's 78A2 (see Figure 4), a single solid state microwave frequency source is used. At this terminal, the baseband signal from the multiplex equipment is first adjusted to the correct level and then fed directly to the FMO (a frequency-modulated microwave source). In a 6-GHz system, for example, the modulated signal usually emerges from the FMO in the region of 1500 MHz, which is then amplified and fed into two multiplier stages. The first stage doubles the frequency and emits a 3000-MHz signal into the second multiplier. The second multiplier again doubles the frequency and sends a 6000-MHz (6-GHz) signal to the antenna. Each time the frequency is multiplied, power is lost. Also, each time a modulated

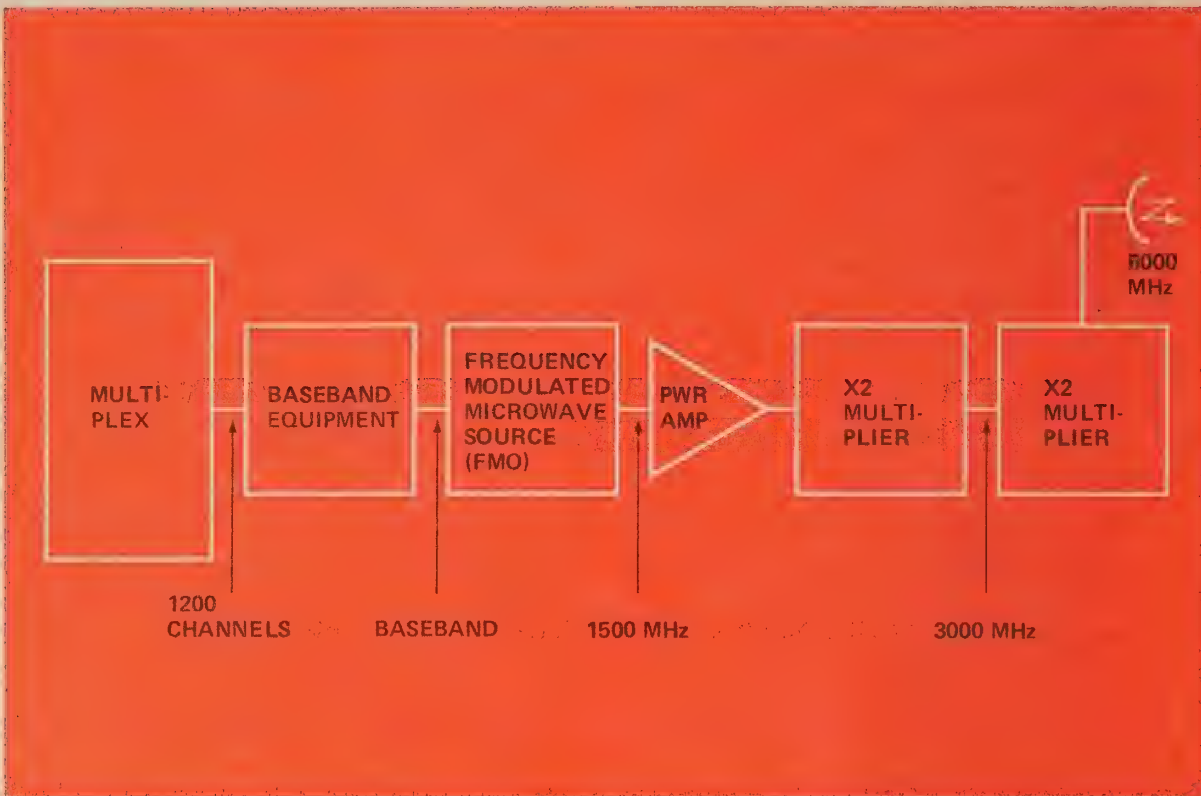


Figure 4. The transmit section of a baseband microwave radio terminal working with telephone multiplex equipment.

signal is multiplied, noise is introduced into the baseband. This is random noise which is inherent in the baseband and which is emphasized during the multiplication process – it is not caused by the multiplier itself.

The great advantage of modern microwave sources and their effect on baseband systems is evidenced in the dramatically few multiplication steps necessary to attain the transmit frequency – two doublers in the case of a 6-GHz system. A previously used microwave source might be capable of modulating at only 150 MHz, instead of the 1500 MHz made possible by solid state sources. *This would mean a 40:1 multiplication ratio would be required compared to the 4:1 ratio for modern microwave sources.* The use of a solid state microwave source results in greater bandwidth, lower noise, better stability, and lower cost.

The Baseband Repeater

Baseband (or remodulating) repeaters are essentially terminal receivers and terminal transmitters connected back-to-back. The function of the baseband repeater is to amplify the incoming signal and relay it to the next repeater or terminal. If, for example, the signal received is 6000 MHz as shown in Figure 5 (6000 MHz referring to the center frequency), it is mixed with a constant 6070-MHz frequency from the microwave source. The difference between these two fre-

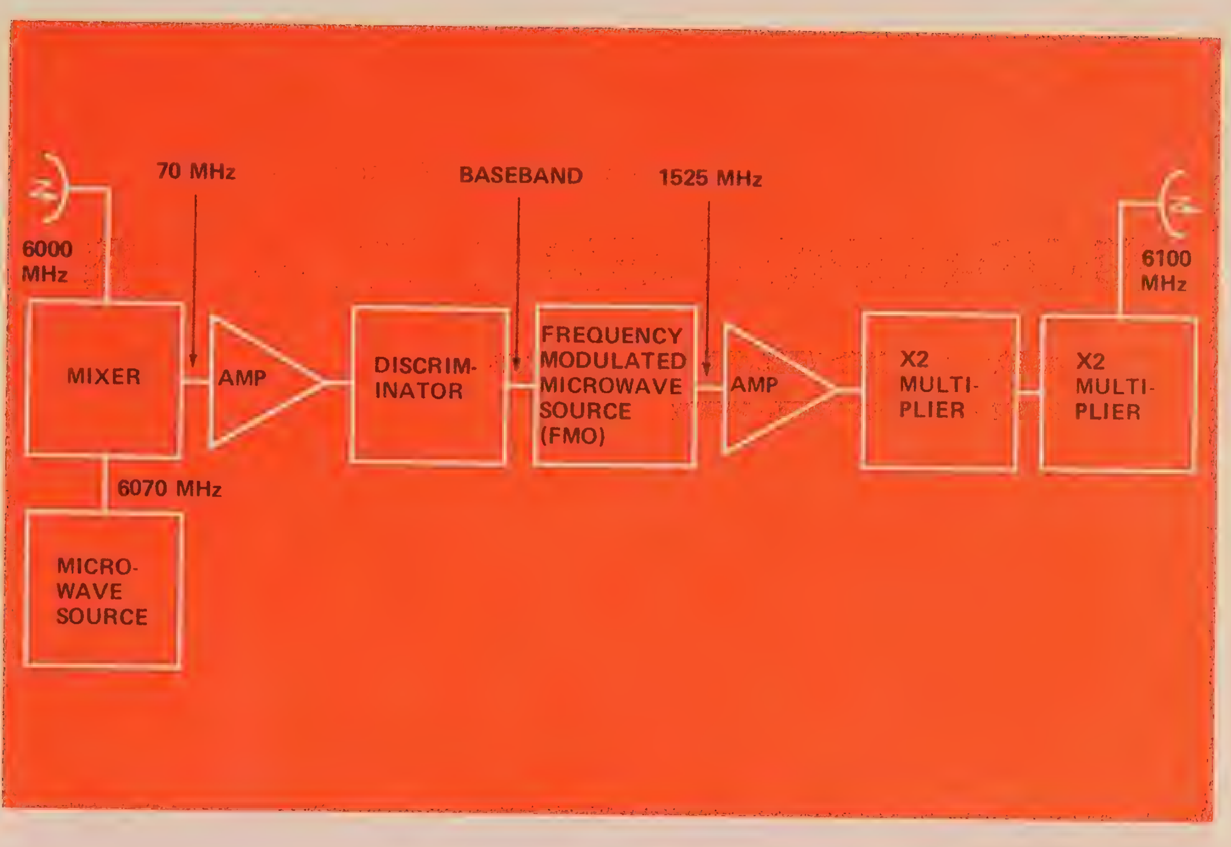


Figure 5. In a baseband repeater, the baseband must be extracted from the RF carrier during the amplification process.

quencies (70 MHz) is then amplified and sent to the discriminator, where the signal is demodulated down to baseband. The baseband now remodulates an FMO, and the output is amplified, and multiplied to the desired 6100 MHz.

The great advantage of a baseband system is that full access to the baseband is available at each repeater point. This availability facilitates the inserting and dropping of baseband information. Furthermore, combining and switching from a working to standby system on a per hop basis is facilitated – affording additional reliability of operation. This, coupled with the high performance of modern baseband systems, provides an ideal arrangement for telephone communications or video networks, for example, which require access to baseband at intermediate repeaters along the route. The improvement in baseband systems can be seen in Figure 6, which shows the typical overall noise performance for a six-hop system (2 terminals and 5 repeaters) of GTE Lenkurt type 78A2 baseband radio. This system is loaded with 1200 channels of SSBSC (single-sideband suppressed carrier) multiplex with a nominal deviation of 141 kHz (rms) per channel, and with –30-dBm RF input to each receiver.

Also, order wire and alarm facilities, for maintenance and trouble locating are almost mandatory at re-

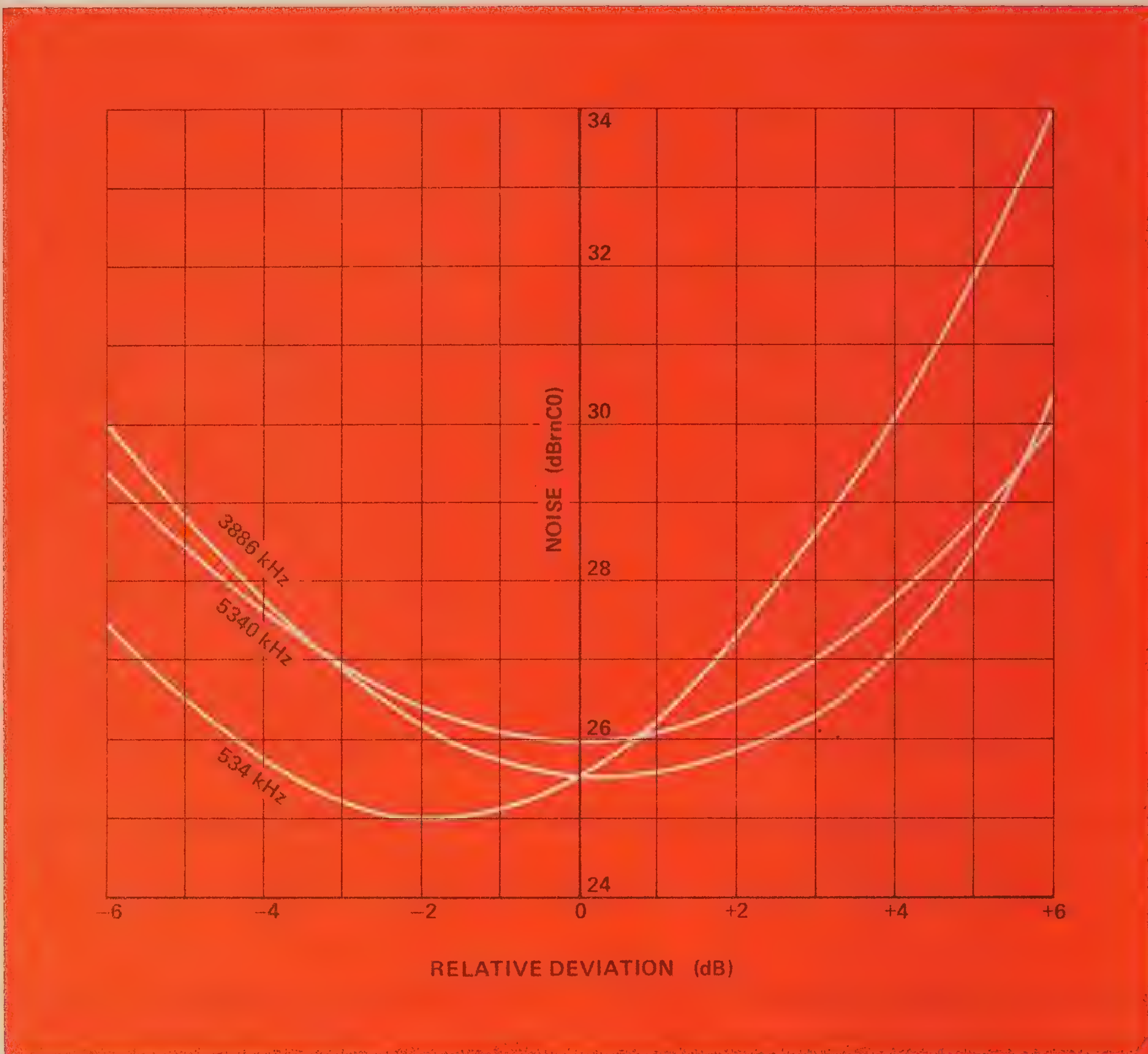


Figure 6. Typical overall noise performance for a GTE Lenkurt type 78A2, 1200-channel baseband radio, in a 6-hop system.

peater stations, and baseband access is required to provide these facilities. This problem is automatically solved with baseband systems.

Receiver Terminals

The process involved in the microwave receiver terminals for either baseband or IF heterodyne operation is the same. The modulated microwave signal from the antenna system, is mixed with a signal from a microwave signal source to produce a "difference" signal at an IF frequency (commonly 70 MHz), as shown in Figure 7. The 70-MHz signal is then amplified and demodulated (by a discriminator) to extract the baseband signal. The baseband equipment adjusts the baseband level and feeds the baseband into the multiplex equipment. The multiplex equipment separates the voice channels contained in the baseband and makes them available for distribution.

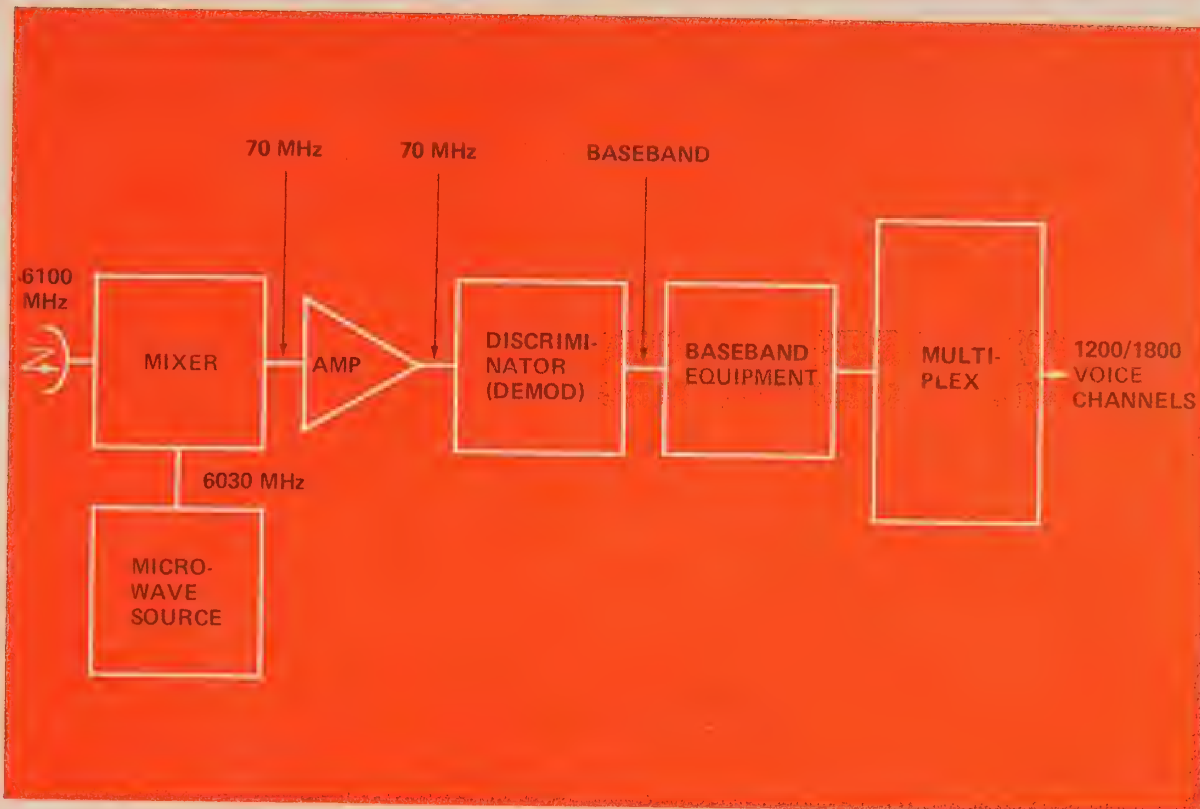


Figure 7. The receiver-terminal function is the same for both baseband and IF heterodyne microwave systems.

It should be remembered that for telephone conversation in the opposite direction, another complement of microwave radio equipment, usually operating through the same antenna system, is necessary.

Baseband microwave systems have undergone many changes since their inception. But the most recent – those connected with new microwave sources and transmission control techniques – have had the most significant impact. This reflects not only a savings to user and customer alike, but also shows continual technological progress in the telecommunications field.

GTE LENKURT

DEMODULATOR

JUNE 1971

12 GHz

a new look by industrials

For industrial users, rain attenuation at 12 GHz may actually be more benign than other transmission outages. So, shouldn't the industrials use the 12-GHz band?

Most industrial users, for a variety of reasons, have usually selected the two lower frequency bands – the 2- and 6-GHz bands. A review of the FCC frequency listings shows only a few hundred licenses in the 12-GHz band, compared to many thousands in the 2- and 6-GHz bands.

The most common reason for this strong preference for the lower frequencies is the susceptibility of the 12-GHz band to rainfall attenuation. Although the effect is present to some degree at the lower frequencies, it increases rapidly with frequency. And, a rainfall intensity causing only a few dB of attenuation at a lower frequency could be sufficient to cause a path outage at 12 GHz (see Figure 1).

Even without the rain effect, users whose operational experience has been in the lower bands tend to prefer them to a band with which they are less familiar. The availability and cost of accessories such as antennas, waveguides, and test equipment have also been an important factor affecting usage.

A New Look?

Several things point toward a "yes" answer to this question. For example, as more microwave systems come into existence, there is growing frequency congestion and in some areas it is already difficult, if not impossible, to find interference-free frequencies for new systems or paths in the 2- or 6-GHz bands.

RF (radio frequency) channels are licensable with 20 MHz of bandwidth in the 12-GHz band; whereas only 10 MHz are available at 6; and 8 MHz at 2, under FCC rules. Thus, of the three bands, 12 GHz is the best suited for wideband services.

Equipment for the 12-GHz frequency band, including antennas, waveguides, and test equipment, is now widely available, with proven quality and reliability quite on a par with equipment for the lower bands. Experience in the 12-GHz band has shown that the only really important propagation differential is that of the rain attenuation effect.

FCC policy for a number of years has been aimed at promoting increased usage of the bands above 10 GHz. One requirement which has been in effect for some years is that any new systems entirely within a municipal or local area must use frequencies above 10 GHz. The Commission has also sought to encourage use of the 12-GHz band for short spur legs on long systems,

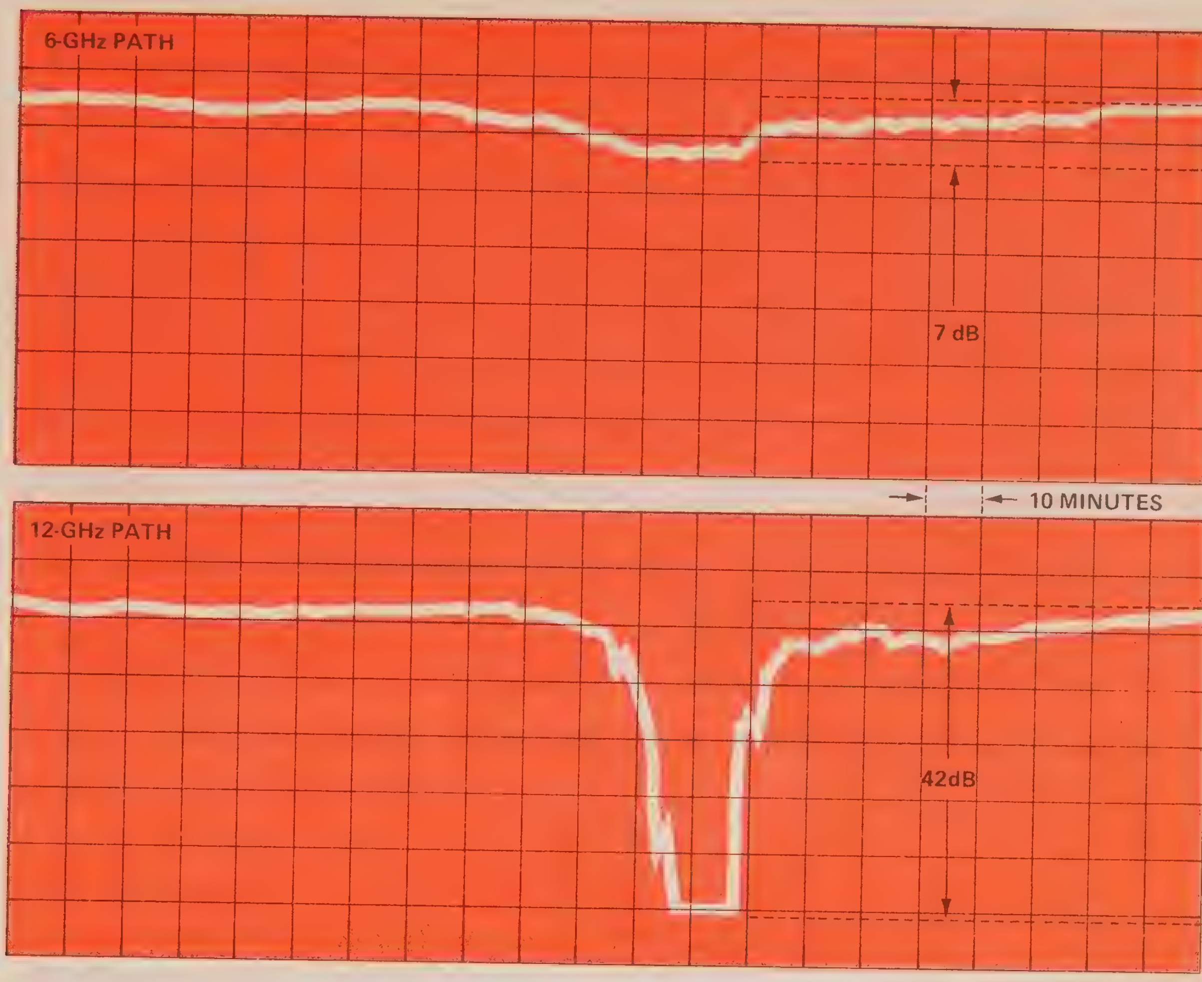


Figure 1. This recording of two simultaneously transmitted channels illustrates the susceptibility of the 12-GHz channels to rain attenuation.

thus keeping the lower frequencies available for backbone routes.

Non-Rain Transmission

Comparing 6- and 12-GHz transmission characteristics under non-rain conditions shows that there is really little difference between the two systems in overall expected performance. For example, on a given path with two given antennas, the path attenuation is greater at 12 GHz than at 6. But, the antenna gain is greater at 12 than at 6. Adding waveguide losses in order to determine end-to-end path loss (path loss plus waveguide losses minus antenna gains) shows the average 12- and 6-GHz systems to be comparable (see Figure 2).

Receiver noise-figures tend to be 1 - 2 dB greater at 12 GHz than at 6 GHz, and for comparable types of equipment there may be from 1 - 3 dB less transmitter output power at 12 GHz than at 6. Thus, with present day equipment, one might expect a given 12-GHz hop to be at a disadvantage of from 2 - 5 dB from the standpoint of equipment. This differential may well be reduced as even better components become available at the higher band.

Path clearance requirements, for a given degree of performance, are slightly lower at 12 GHz than at 6 because the Fresnel zone radius at 12 is only about three-quarters as large as at 6 (see Figure 3). However, this difference is not enough to be signifi-

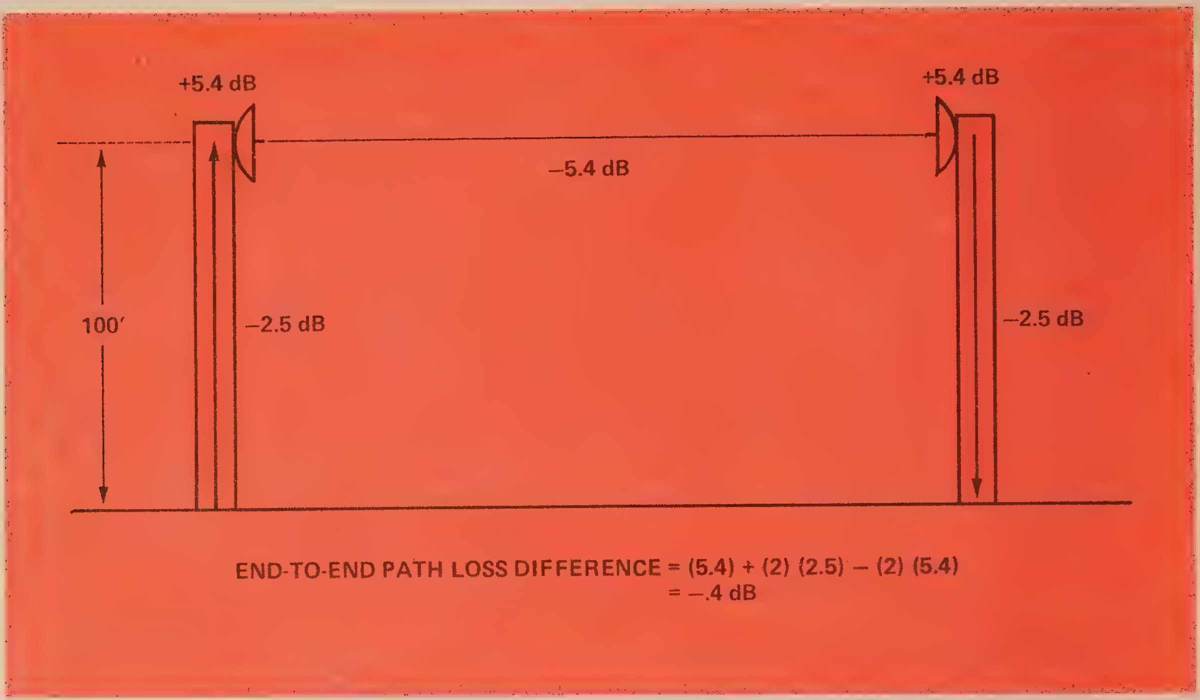


Figure 2. When compared to a 6-GHz system, a 12-GHz system has a similar end-to-end path loss.

cant. Fading of the multipath type is quite similar in nature to that experienced at 6, though it is now generally considered that the fading at the higher frequencies is somewhat greater. Where necessary, space diversity can be used to overcome multipath fading, and the same spacing is even more effective at 12 than at 6 GHz.

Summing up, the total net differential between the 6- and 12-GHz bands is relatively small, and in some cases may even favor the 12-GHz band. Thus, if it were not for the rain attenuation effect, there would be no reason why the 12-GHz band could not be used in much the same way as the lower frequency bands.

Rain Attenuation

Rain attenuation at the higher microwave frequencies has been under study and investigation for more than 25 years. Much is known about the qualitative aspects, but the problems faced by the microwave transmission engineer — that of making quantitative estimates of the probability distribution of the rainfall attenuation for a given frequency band as a function of path length and geographic area — remains an extremely difficult one.

In order to estimate this probability distribution, instantaneous rainfall data is needed. Unfortunately the available rainfall data is usually in the form of a statistical description of the amount of rain which falls at a given measurement point over various time periods — generally at least an hour in length.

The rain-induced attenuation along a given path at a given instant in time, however, is a function of the integrated effect of the rainfall existing at all points along the path and is affected not only by the total amount of water in the path at that instant but also by its distribution along the path in volume and drop size.

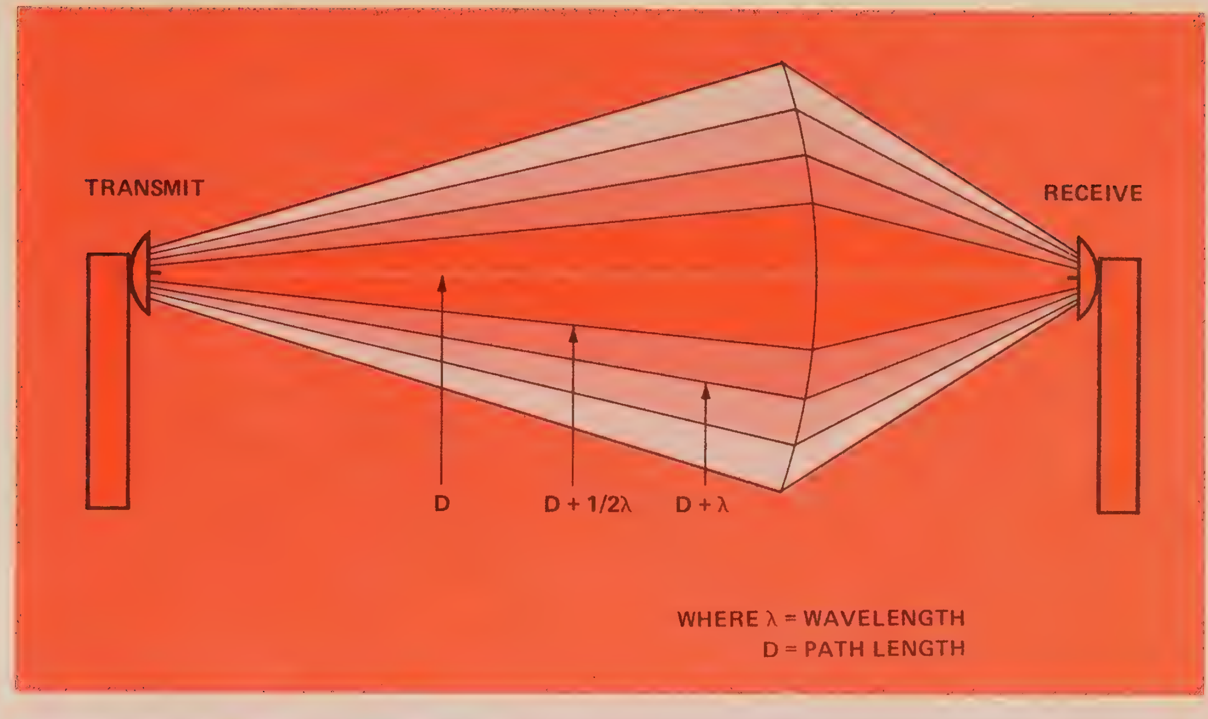


Figure 3. Fresnel zone radii are a function of the signal wavelength, and consequently, the signal frequency.

For heavy rain rates the instantaneous distribution of volume and drop size along the path is highly variable, and is difficult to predict with any sort of accuracy from the kind of rainfall data generally available.

One of the earliest and most comprehensive attempts at developing a workable prediction method was carried out by Bell Laboratories in the 1950's, and was described in a classic paper by Hathaway and Evans (1958)*. In their paper Hathaway and Evans developed a method of predicting annual outages for microwave paths operating in the 11-GHz common carrier band, as a function of path length, fade margin, and geographical area within the contiguous United States.

This study has proved to be a worthwhile prediction tool, and when used with a recognition of its limitations, is still one of the best references available for microwave engineers working within the United States. The Hathaway and Evans method can be modified slightly to adapt it to the 12-GHz industrial band rather than the 11-GHz common carrier band (see Figures 4 and 5).

Increasing fade margin and shortening path lengths are the most readily available tools for reducing the per hop outage in a given area. For fade margins other than 40 dB as shown in Figure 5, correction factors (shown in Figure 6) can be used.

The total annual rainfall in an area has almost no relation to the rain attenuation for the area. Within the U.S., the northwestern states, for example, have the greatest annual rain-

*Other investigations were carried out for the Federal Aviation Administration and are covered by Report No. FAA-RD-70-21, *Rain Attenuation Study for 15-GHz Relay Design*, and Report No. FAA-RD-70-47, *Weather Effects on Approach and Landing Systems.*

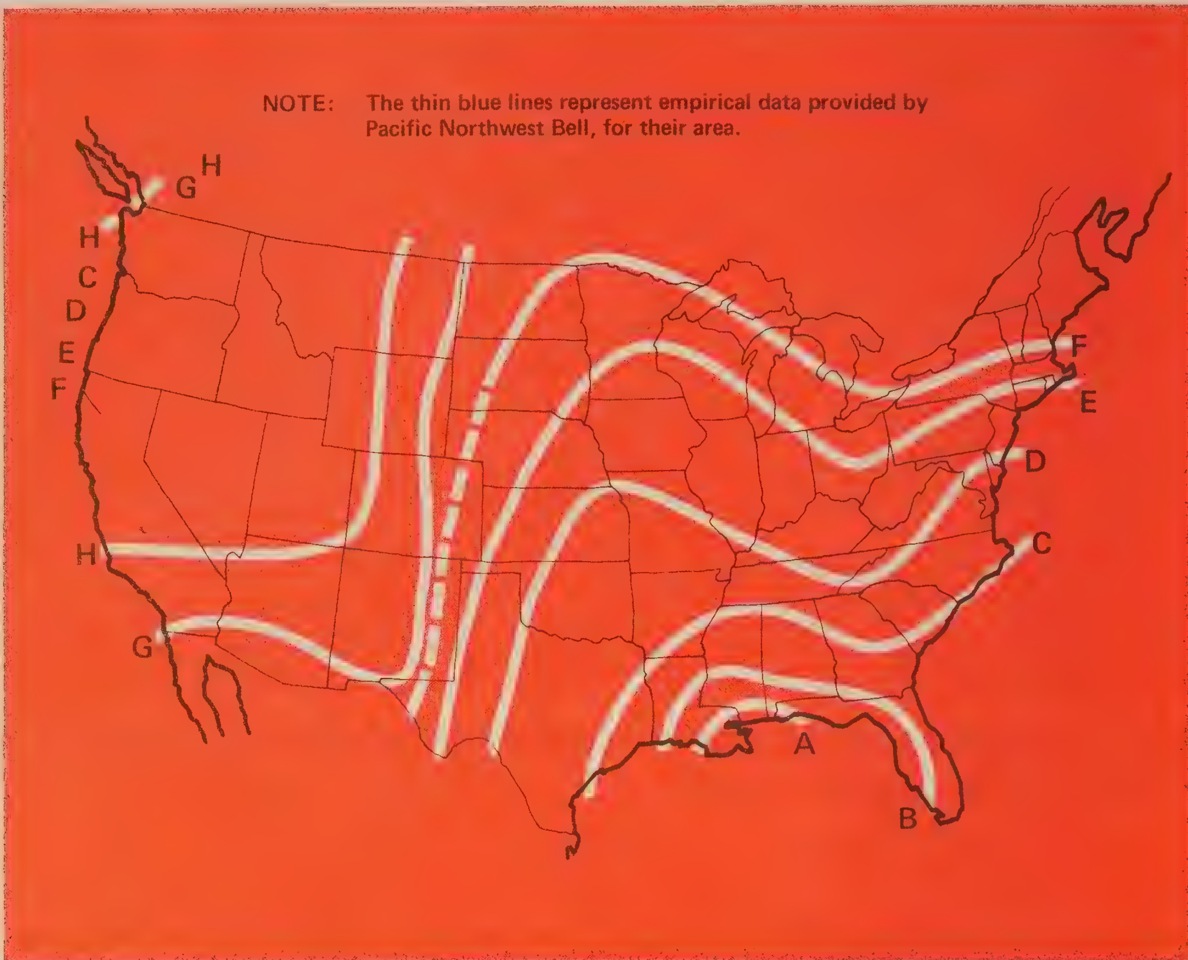


Figure 4. The contours of this map have the same average rainfall distribution and can be used with Figure 5 to predict the effect of rainfall on outage time.

fall, in excess of 100 inches per year, but it is produced by long periods of steady rain of relatively low intensity at any given time. Other areas of the country with much lower annual rates experience types of rainfall such as thunderstorms and frontal squalls which produce short duration rains of extreme intensity, and it is the incidence of rainstorms of this type which determines the rain attenuation characteristics of an area.

Even the rain statistics for a day or an hour have little relationship to the excess path attenuation. A day with only a fraction of an inch of total rainfall may have a path outage due to a short period of extremely high intensity, while another day with several inches of total rainfall may experience little or no excess path attenuation because the rain is spread over a long time period.

Reliability Objectives

A company with its own communications system is the end user as well as the operator of the system, and thus is in a more flexible position than the common carriers who are selling communications service to the public. The private user can meet exact reliability requirements for different parts of a system. For example, a spur leg to a facility of minor importance might be considered satisfactory with a pre-

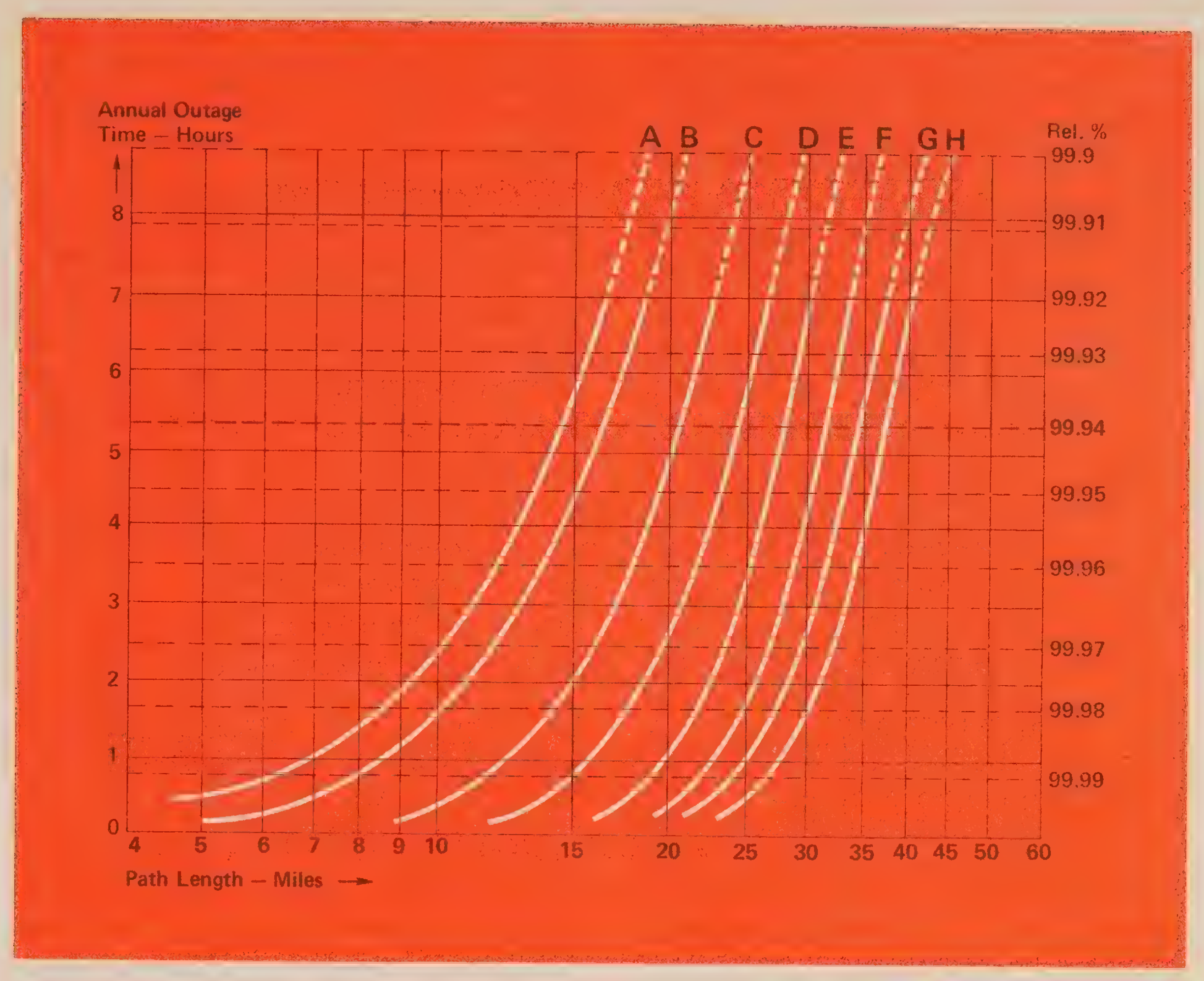


Figure 5. These curves, for use with the contour map of Figure 4, are based on 12-GHz paths with 40-dB fade margins.

dicted outage of several hours per year, in sharp contrast to the requirements for a backbone hop of a long system or into a site of major importance.

Fade Margin	Change annual outage by:
35 dB	+ 20%
45 dB	– 15%
50 dB	– 25%

Figure 6. Using these correction factors, Figure 5 can be adapted for paths with different fade margins.

In considering how to establish realistic outage or reliability objectives, several things need to be kept in mind. A single overall design objective for not more than X hours, minutes, or seconds outage over some period such as a year, is an over-simplification. The character of the particular kind of outage and its effect on the system should be taken into account and perhaps there should even be different objectives for different types of outage.

For example, propagation outages due to multipath fading are usually short. An outage of an hour per year due to multipath fading might represent 1,000 or more individual outages,

averaging about 3 or 4 seconds each. On the other hand, propagation outages totalling an hour per hop due to rain attenuation, on a path with a large fade margin, might consist of four or five individual outages averaging ten to fifteen minutes each. The effects of these two types of system outage would be quite different in nature.

A distinction should be made between communications circuits for which an outage of a few seconds or a few minutes is just a nuisance or an inconvenience, and circuits for which such an outage might result in danger to life, great economic loss, or other catastropic consequences. The suitability or unsuitability of a rain-affected band such as 12 GHz could differ widely for these two situations.

Even if the maximum possible reliability objectives are established and a path or a system is engineered to the full limit of the state of the art, the possiblity of an outage can never be eliminated but can only be reduced to a very low probability. Thus it is imperative to make any ultra-important services as fail-safe as possible against a loss of the communications channel. Therefore, regardless of the degree of reliability, a system should be engineered so that if an outage does occur it can be tolerated or its effects at least kept within reasonable bounds.

It seems that in some cases, perhaps many cases, a somewhat more relaxed attitude might be taken toward rain-induced outages than toward multipath outages or even equipment outages. In several respects such rain outages seem to be somewhat benign in nature. If the fade margins are kept high and the paths are not stretched out too much, even in the less advantageous areas of the country, the number of outages per year should not be very large, and the length of individual outages on a hop should only rarely exceed some two to perhaps twenty minutes.

Furthermore, such outages would occur only with extremely heavy rainfall somewhere along the path, and the conditions when this is likely to occur are usually known in advance and fairly well publicized. This type of outage should be considered tolerable since it occurs rather infrequently, seldom happens without some advance warning, doesn't last long when it does happen, and is self-healing.

For high reliability systems, usually involving long-haul systems with a great many hops in tandem, the per hop objectives may be as stringent as 99.9999% or so, allowing only about 30 seconds outage per year. Short haul systems, up to say ten hops, might have per hop design objectives of about 99.999%, roughly 5 minutes outage per year. Spur legs or single hop systems may be designed for something on the order of 99.99% or about 53 minutes outage per year. Objectives of this kind are typical of those used in the telephone industry, for public service networks. For other situations, and for other types of service, even lower reliabilities may be acceptable, down to 99.9% or about 9 hours outage per year.

Figure 5 shows the predicted annual outages as percentages, and a little study of the numbers indicates that even in favorable areas of the country one would have to use quite short paths in order to get much beyond the 99.99% line (53 minutes per year). Attempts to extrapolate down to the 99.999% or the 99.9999% areas would be subject to great uncertainty.

Using Figure 5 as a guideline, it is apparent that there are few areas where it would be feasible to use 12

GHz as a part of the backbone route of a long-haul system using conventional path lengths, particularly one where requirements are high. On the basis of this data, it seems that the 12-GHz band would be most useful for situations in which the per hop reliability design objectives fall in the range of 99.9% up to at least 99.99%.

Diversity Plans

In the adjacent 11-GHz band, widespread use has been made of cross-band diversity systems, a form of frequency diversity in which one of the frequencies is in the 6-GHz common carrier band and the other in the 11-GHz common carrier band. This combination is a very effective one and can be used in any part of the country without any particular concern about path length. In heavy rain areas the 11-GHz half of the path can be expected to experience outages during heavy rainfall but the 6-GHz path will be only slightly affected. Furthermore, multipath fading is seldom experienced during heavy rainfall so the temporary outage of the 11-GHz path will have no effect on the system unless the 6-GHz path fails during this time because of equipment trouble, a very low probability situation in well engineered systems.

Cross-band diversity between the 6-GHz and 12-GHz industrial band is technically feasible and would be equally useful. But under FCC licensing policies in the industrial bands, cross-band diversity is only available in special cases where a definite need can be demonstrated. Also its use would not result in any reduction in the 6-GHz band usage, since industrial users are not allowed to use in-band frequency diversity.

There are a few cross-band hops in the industrial band, and Figure 1 is a recording from one such path, showing an example of a rain outage on the 12-GHz half of the path. This is quite typical of the rain induced outages in the higher bands, both in depth and length. The excess attenuation of the 12-GHz side exceeded 40 dB for a period of about 12 minutes, bottoming the recorder. The actual depth may have been considerably greater than 40 dB. During the same interval, the 6.7-GHz path experienced only a small amount of rain attenuation – about 7 dB at the maximum and of no real significance to the system.

This particular path is about 21 miles long, but the appearance of the attenuation event indicates that it was probably caused by a single rain cell occupying a relatively small portion of the path, rather than by uniform rain spread all along it. This leads to the conclusion that the event might have occurred, in about the same magnitude, even if the path had been quite short, perhaps even as short as five miles. However, the number of such events which would be expected to occur over a given period of time in a five mile path would be only one-fourth the number which would be expected to occur in a twenty mile path. In other words, it seems likely that for paths over about five miles it is not the amount of excess attenuation which determines path length, but rather the number of expected outage events and the total length of the expected outages.

Another approach to a high-reliability system uses a combination of very short path lengths plus route diversity to defeat the rain attenuation problem. These path lengths are from 2 to 5 miles, so the number of repeaters would be large compared to conventional microwave systems where the hops average something like 25 - 30

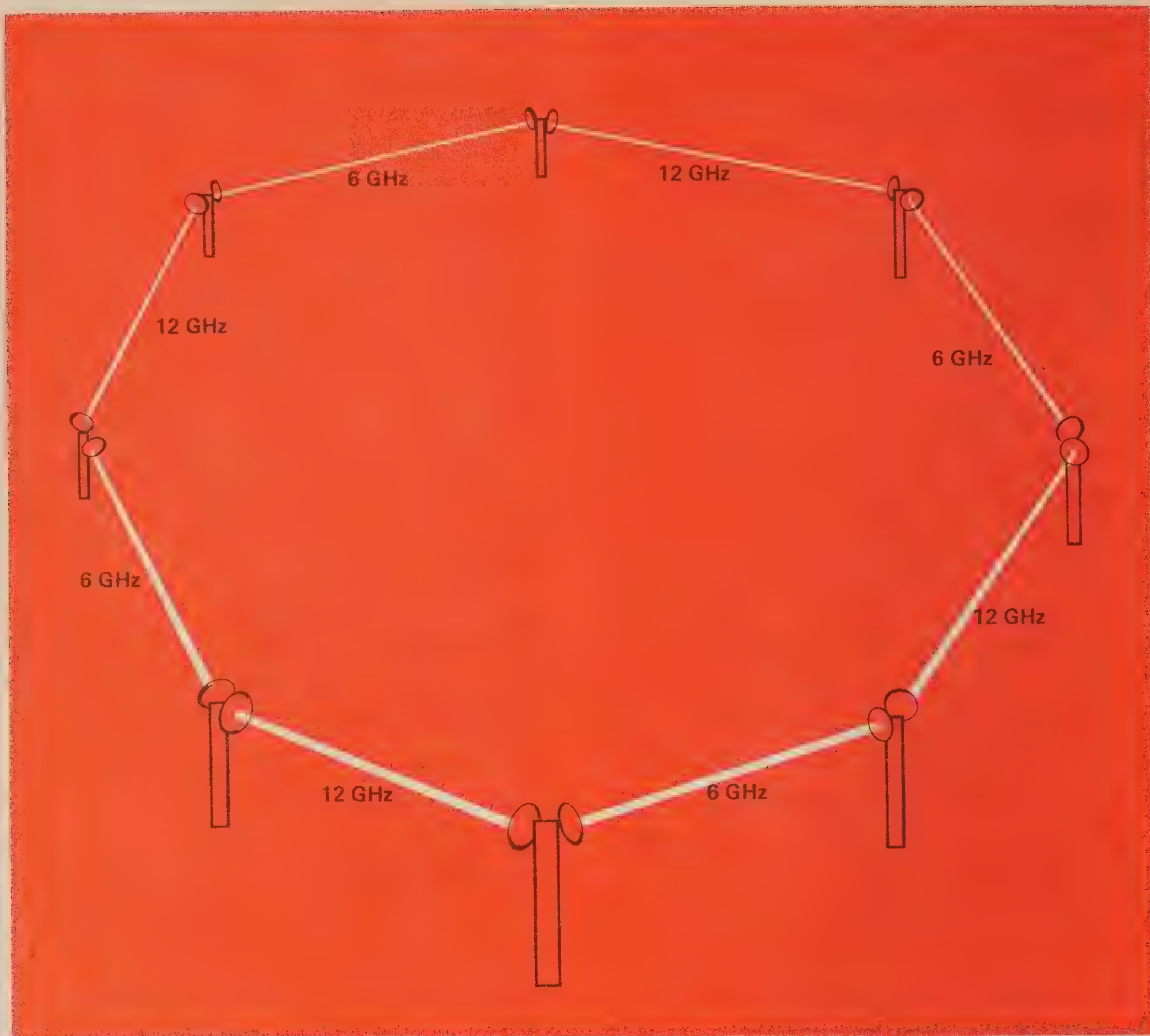


Figure 7. If one link of a closed loop network fails, a loop diversity arrangement will reverse the transmission direction to complete the transmission path between terminals.

miles in length. A system of this kind is called a "pole line radio." For such an arrangement, repeaters must be low in cost, highly reliable, virtually maintenance-free, low in power consumption, and capable of being placed in an inconspicuous housing at the top of some sort of simple pole structure. In addition the repeaters must be broadband to allow high channel density and each repeater must introduce a low amount of distortion and noise. The latter effect would be achieved by using PCM transmission for the system, with regeneration at sufficient intervals along the line to keep the error rate at the desired low level.

This whole technique, though promising, is still in the experimental stage and its potentialities and problems are still largely unknown. One basic principle is that it would require new frequency bands, not already in use by systems with conventional techniques, since the integration problems between the old and new would be large.

For the industrial user there is little likelihood that systems of this type would come into use in the 12-GHz

band, though eventually, there may be something of the kind at a higher frequency.

However, there is one special situation in which a form of route diversity is already being used by some of the industrials. This is the so-called "loop system" in which a microwave system is laid out in the form of a closed loop with some sort of automatic sensing and switching system (see Figure 7). With such a system an outage or failure in any link will reverse the transmission direction around the loop.

Since the evidence indicates that intense rainfall at a given instant is unlikely to occur at two widely separated points, it seems that in such a system it would be quite feasible to use 12-GHz and 6-GHz frequencies alternately around the loop. With no two 12-GHz paths adjacent to one another, the physical separation would seem more than adequate to insure that rain outages would be essentially non-correlated and the likelihood of two paths being out simultaneously would be small.

Unfortunately, most communications systems proceed in more or less a straight line and are thus not adaptable to loop systems, so the possible application is somewhat limited. One possible application would be for systems connecting off-shore drilling platforms, which often are located in a manner suitable for looping. Another would be power companies, whose grid-type power transmission networks tend to be arranged in natural loops.

It should be kept in mind that loop systems, though extremely useful against outages which are relatively rare and rather lengthy, do not provide any protection against rapid, short fading of the multipath type. Where this type fading exists, per hop space diversity should be used to combat it, even with a loop system.

Gain from Experience

The adjacent higher frequency band, from 12.7 - 13.5 GHz, assigned to television auxiliary stations, has had rather widespread usage, a good part of it in microwave relay systems serving CATV systems. For this type of service somewhat lower reliabilities are deemed acceptable, and the systems are generally engineered with little or no regard for possible rain outage, even in the less favorable areas.

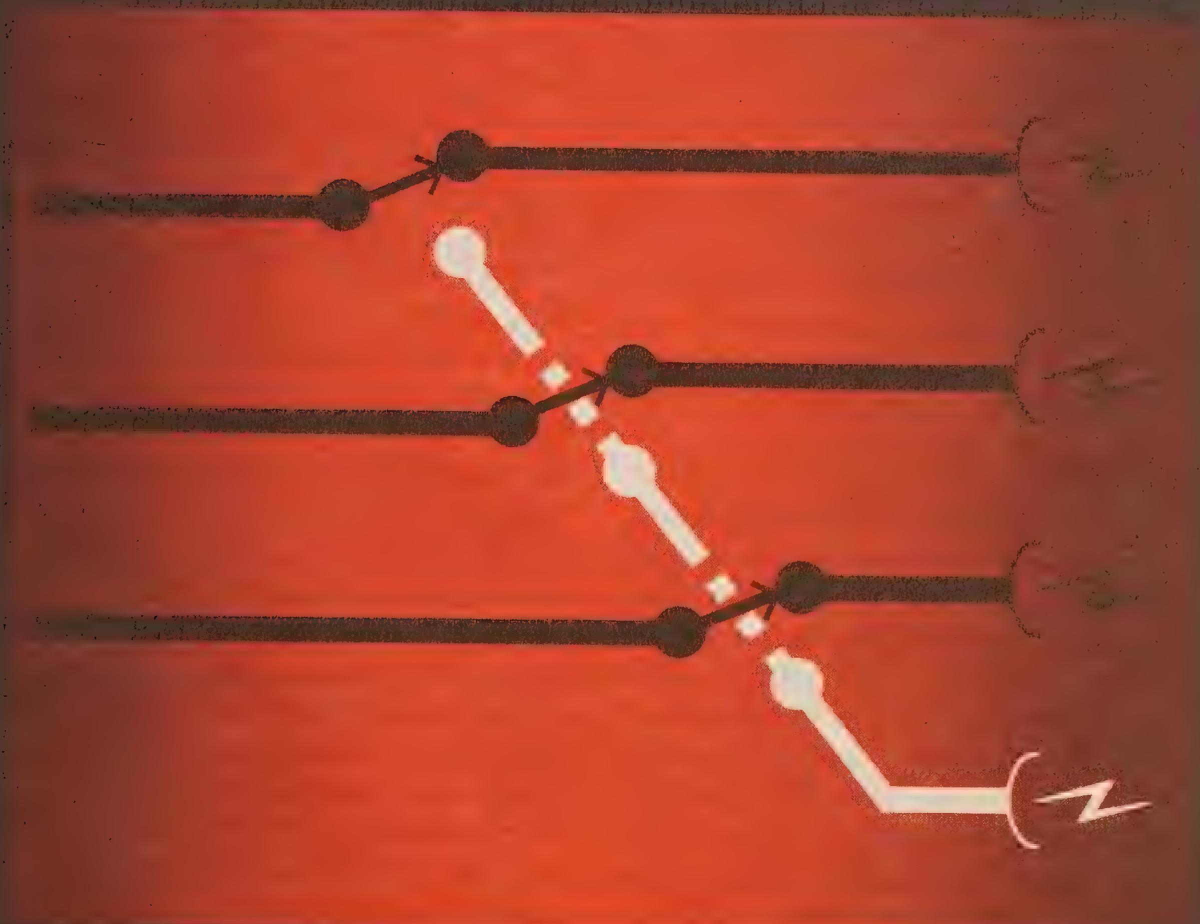
In some areas the common carriers and some of the industrial users have found themselves pushed into the higher frequency bands by congestion in the lower bands, and have generally found that although considered totally unsatisfactory before, the higher frequency bands can be quite useful.

GTE LENKURT

DEMODULATOR

JUNE 1972

Multiline Protection Switching



A multiline switching system provides protection for microwave transmission paths. It is not only an economical method of protection, but is also a means of conserving bandwidth in crowded microwave bands.

Protection of Microwave communications channels is necessary in the event a working channel experiences an outage due to equipment malfunction, high noise, or severe fading. Frequency and space diversity and multiline switching are the most commonly used methods of microwave system protection. Both frequency and space diversity have the expensive disadvantage of requiring duplication of equipment (transmitters, receivers, and antennas) to protect a single radio channel. In addition, the extra frequency spectrum necessary for frequency diversity may not be available in many areas.

Multiline Switching

The concept of multiline switching is to allow one or two microwave radio channels to protect several radio channels. Multiline switching systems, such as GTE Lenkurt's 757C, utilize one or two radio channels to protect up to six working microwave radio channels. Such a system is effective on any high density radio route.

A typical section of a microwave radio route which uses multiline switching consists of about one to ten hops, with six hops as the average. This six-hop distance corresponds to approximately 150-180 miles, depending on the terrain and the operating frequency of the system. Greater distances are obtainable by installing additional switching systems in tandem. Thus, a microwave radio route of from a few hundred to several thousand miles may be divided into a specified number of switching sections, each containing the necessary hops to span the required distance with the greatest reliability.

The greatest advantages of a multiline switching system are economy in capital and efficient use of available frequency spectrum. Protection for six radio channels requires only one or two additional radio channels. An additional economic factor is in the flexibility of the system. A user of telecommunications equipment, for example, whose present requirements call for only one microwave radio channel, may install a multiline switching system. Initially, this will give him a one-for-one protection system. As his requirements for radio channels increase, the user can easily expand his facility and still retain adequate channel protection. This type of expansion allows the user to go from one-for-one

Figure 1. Switching centers, such as General Telephone's Clearwater, Florida office, may utilize a variety of baseband, IF, and modem multiline switches.

protection to two-for-six protection (two protection channels for six working channels).

Multiline Applications

A typical high density radio route contains several microwave radio channels. Each radio channel may in turn carry hundreds of information channels consisting of voice, data, or video and program channels. At present, as many as 1800 two-way FDM (frequency division multiplex) channels may be transmitted on a single radio channel. An outage on one such radio channel for a prolonged period would have wide-ranging consequences affecting a great number of customers. Multiline switching provides economical and efficient protection for such a system.

A multiline switching system is not restricted by any particular type of radio transmission equipment. It can be made applicable to relatively short routes which use baseband remodulating microwave radio, or long-haul IF interconnected heterodyne radio (see Figure 1).

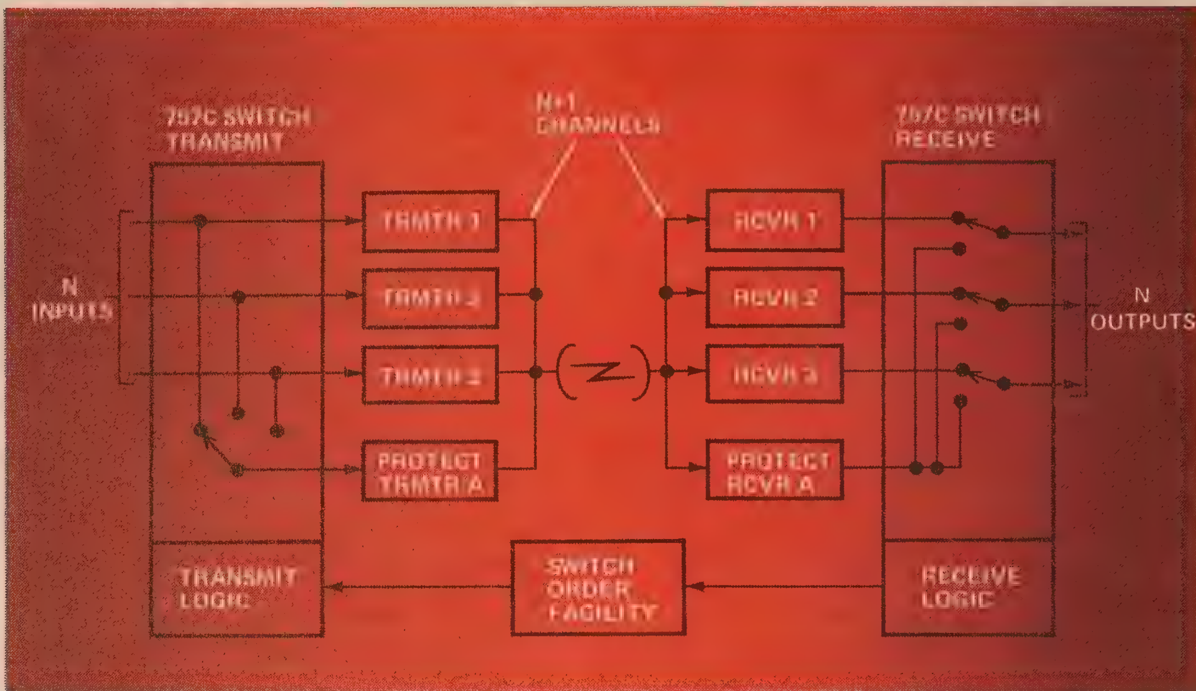


Figure 2. In a 1-for-3 protection system, one channel protects three working radio channels. Here, channel A protects channels 1, 2, and 3.

High density multiline switching is also used with coaxial cable systems, and although the switching philosophy is the same as for radio, signal loss considerations along the line are different. Only microwave radio considerations will be taken into account in this discussion.

Function of A Multiline Switch

The general configuration for a one-for-three multiline protection system going in one direction is shown in Figure 2. The mirror image of this system would appear for transmission in the opposite direction.

The basic operation of a multiline switching system is to transfer information from a failed circuit to a protection circuit, thereby bypassing the fault. There are various methods of doing this, but the basic philosophy is the same for most systems. On a given multiline-switching transmission path such as shown in Figure 2, there is a terminal of equipment which has N + 1 outputs on the transmitting end; N being the number of working channels in a system. On the transmit side of the switch, the common connection goes to the protection channel. Each of the poles on that switch, bridge the working radio channels. At the receive end, each working path has a two position switch. It is important that the switches on both the transmit and receive ends remain in constant synchronization. This synchronization is controlled by the logic contained in each end of the multiline switching section.

Types of Multiline Switches

There are two types of multiline switches — baseband and IF — each corresponding to the type of radio transmission used in a particular microwave system. Basically, the only significant difference between baseband and IF (intermediate frequency) systems is in the transmission-path equipment.

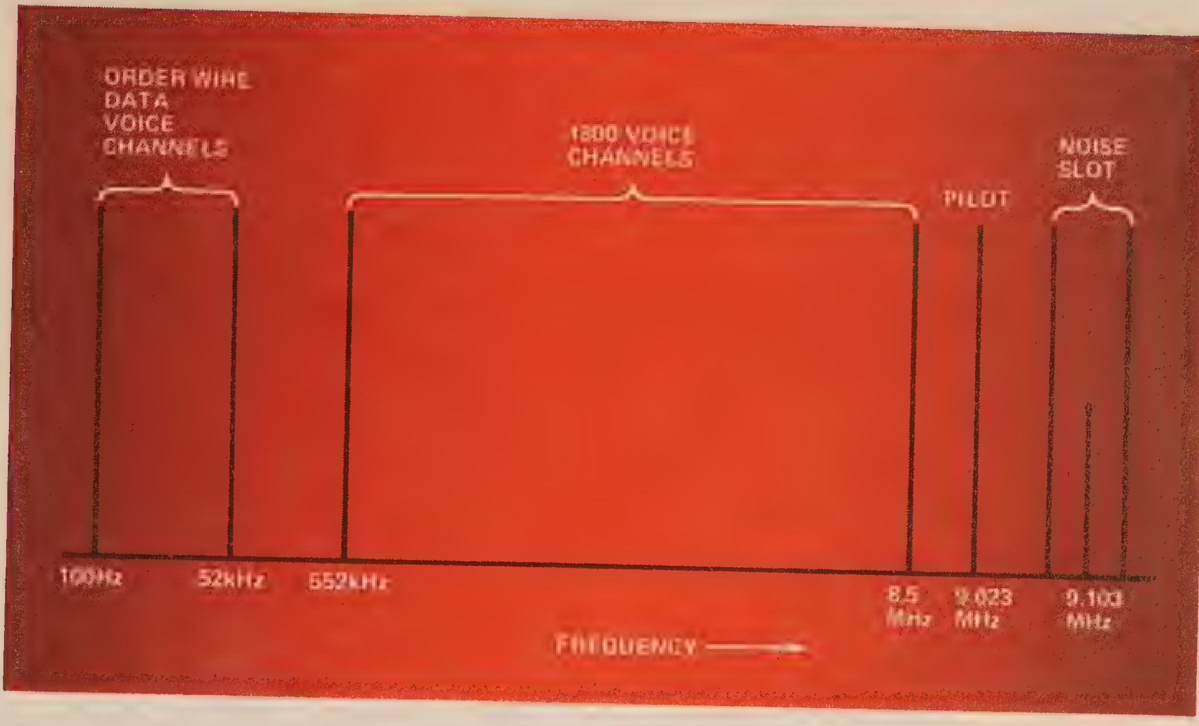


Figure 3. Example of frequency distribution in a microwave radio channel with an 1800 voice-channel capacity.

For a switching system to operate, it is necessary to have a way of determining when a transmission path has failed. This is done in the GTE Lenkurt 757C by monitoring noise and continuity. In baseband and some IF systems, each radio channel contains a fixed-frequency continuity pilot and a noise slot which are continuously monitored by the switching system. Both the pilot and noise slot are generally above the occupied baseband spectrum as shown in Figure 3. In IF systems which do not utilize a continuity pilot, a fixed noise slot and the intermediate frequency power at 70 MHz are monitored by the switching system. If a power-level change occurs at 70 MHz, or if increased noise appears in the noise slot, the switching system will initiate a transfer.

Switching Thresholds

The noise threshold of a microwave radio system is that point at which noise becomes great enough to render a channel useless for information transmission. In a multiline switching system, the "switching threshold" is not necessarily the same as the noise threshold of the radio equipment. The switching threshold – the point at which a particular switching system is designed to initiate a switching sequence – may be a few dB different than the noise threshold of the radio equipment, depending on the requirements of the system. These requirements will vary with the nature of the information being transmitted. When a switching threshold is surpassed, the system will initiate a switch to the protection channel.

Excessive noise in a microwave system may be caused by equipment failure, atmospheric changes which cause transmission-path fading, or adverse weather conditions such as heavy rainfall or fog. These conditions are detected by the multiline switching system by monitoring the noise slot. A degeneration of the pilot level may

TYPE OF TRAFFIC	MAXIMUM TOLERABLE INTERRUPT TIME	EFFECT IF TOLERANCE IS EXCEEDED
VIDEO	< 100 MICROSECONDS	LOSS OF SYNCHRONIZATION ROLLING
TELEGRAPH (50 BAUD)	< 1 MILLISECOND	ERROR
DATA	< 10 MICROSECONDS	ERROR
MESSAGE CIRCUITS	< 100 MILLISECONDS	SEIZURE OF CENTRAL OFFICE SWITCHING EQUIPMENT

Figure 4. The effect of an outage depends on the type of information being transmitted.

be caused by circuit interruption due to man-made error, a component failure which will not result in excessive noise but only in a total quieting of the transmission path, or a gradual deterioration of the circuit due to a long-term component failure. The occurence of any one or all of these conditions is sufficient to initiate a transfer sequence to the protection channel.

Multiline Switch Operation

In describing logic operations, it is often convenient to use anthropomorphic terms. For example, in achieving synchronization between the transmit and receive sections of a multiline switching system, the transmit and receive logic of the system must be able to "talk" to each other. The transmit logic must function under the control of the receive logic so that synchronization may exist between both ends.

When the pilot or noise detectors recognize a failed channel, the receive logic is notified. The receive logic initiates the switching sequence, and via the switch order facility, notifies the transmit logic to bridge the traffic of the failed channel to the protection channel. The switch order facility carries information from the receive logic to the transmit logic, either by transmitting the command information of the receive logic to the transmit logic using an unused portion of frequency spectrum of one or two radio channels going in the opposite direction, or by external order wire radio or cable facilities. The transmit logic must then determine the availability of the protection channel. Is the protection channel available? Is it already being used to protect another channel? These are some of the decisions which the transmit logic must make before initiating a bridge to the protection channel. If the channel is already in protection mode, or if it is not functioning properly, it will not react to the switch order request. If the protec-

tion channel is available, the transmit logic then bridges the traffic of the failed channel to the protection channel which then relays the information to the receive end.

The receive end detects that the information of the failed channel is now on the protection channel. This will tell the receive logic that the transmit end has bridged the correct channel to the protection channel and that the protection channel is operating properly. The receive logic then completes the switching sequence by transferring the information of the failed channel around the fault and on to its original route.

The time required for the complete switching sequence is no more than 30 milliseconds, depending in part on such things as the number of repeaters between terminals and type of switch order facility used. The *actual* interrupt time of the circuit is approximately two microseconds since the switching sequence begins before a channel has completely failed. The switching sequence begins at a point predetermined by the switching threshold of the system.

Effects of Outages

The effects of an outage on the various types of information is unique to the information itself. For example, message circuits may sustain much longer outages or "hits," without errors, than can a data circuit. The effects of hits on various types of information carried by a radio channel which is protected by multiline switching is shown in Figure 4.

Other Uses

The protection channel of a multiline switching system may serve as a restoration circuit in times of emergency to re-establish communication service when disruption of radio channels not within its own system has occurred. Also, the user may use the protection channel to transmit low-priority information which could be immediately dropped should protection be necessary for one of the main working channels of the system. In this way, a user may get the maximum benefit from his multiline switching system.

Priority Channel

Multiline switching systems may offer a priority-channel option. A priority channel is that working channel which is regarded by the user as the most important. Should this channel fail, the protection channel will immediately take its place, even if it is protecting another channel at that time.

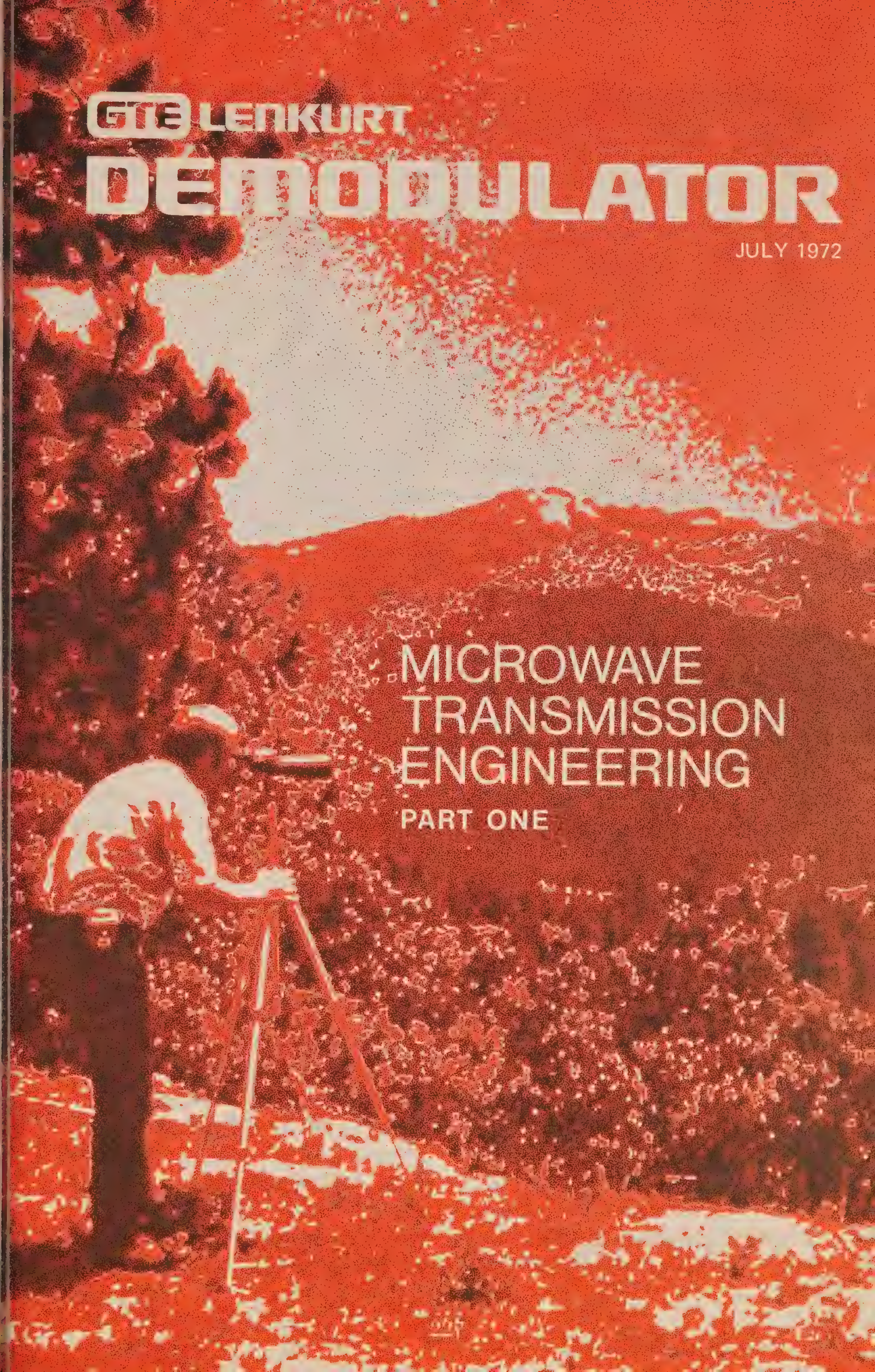
Many telephone companies which use multiline switching do not utilize the priority option because they consider any working channel just as important as any other working channel. In this case, no channel takes precedence.

A multiline switching system is an economical way of protecting the channels in a microwave or coaxial system. It is not the answer to every system-protection problem, but it is certainly an efficient and reliable method of providing protection for high density systems.

GTE LENKURT

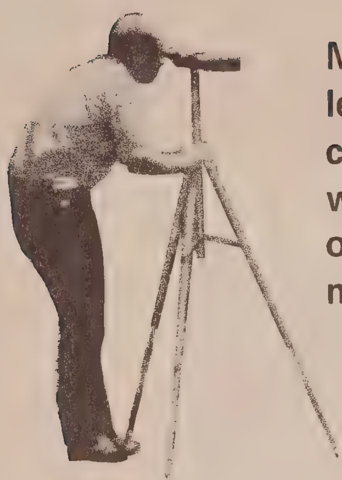
DEMODULATOR

JULY 1972



MICROWAVE TRANSMISSION ENGINEERING

PART ONE



Microwave transmission engineering requires collecting large quantities of data and involves a precise study of the terrain along the proposed microwave path. Interpretation of this information, based on theory and practice, provides the basis for final microwave system design.

The basic purpose in engineering a microwave path is to achieve a path which will meet the requirements for long term median noise, and which will also ensure that outages due to fading below a predetermined level are minimized. In general, and particularly in high-density systems, it is necessary to consider interference noise, waveguide and feeder distortion, and basic thermal and intermodulation noise.

The microwave transmission engineer has to work with many different factors, some of which interact with each other, in order to come up with solutions which are satisfactory both technically and economically. Before starting on any actual path layouts some comprehensive decisions will have to be made as to what the microwave system must do, such as the number of channels for present and future needs, type of noise performance necessary, and kind of reliability required. The decisions may even be more detailed, for example, choice of particular frequency band and type of protection arrangement may be specified. Some of these decisions may come as a result of preliminary path study; however, in any event, they must all be firmed up before final path survey work is undertaken. Figure 1 illustrates some of the "tools" used for preliminary path studies.

Preliminary Planning

The first practical aspect of laying out a system is to know the points the system will connect. The simplest system would be a one hop system, from Town A to Town B 25 miles away, in flat country where a line-of-sight path can be easily achieved. If the distances involved are greater, or the intervening terrain unfavorable, it may be necessary to use more than one hop. The system would then require the use of repeaters.

In relatively flat country, a practical rule of thumb is that repeater spacings are generally limited to about 25 or 30 miles, unless extremely high towers are practical. On long systems where several repeaters are needed, and particularly where a number of points are to be served by the system, a number of choices of route may be

Figure 1. Some of the common tools used in preliminary path surveys include drafting instruments, topographical maps, equivalent earth profile templates, and preliminary profile charts.

available, and considerable study may be needed to select a favorable one. Figure 2 shows a system layout with the necessary repeaters.

Regardless of the system size, for proper system operation, it is necessary for each path to have adequate clearance under all expected atmospheric conditions. To determine clearances, the actual topography of the path and also the height and location of any obstacles along the path, such as buildings and trees, must be known. The topography of an area can be studied prior to going out in the field through the use of topographical maps.

In most cases it will be found that a thorough map study will narrow the problem down considerably, particularly in the case of multi-repeatered systems with a wide range of path choices. By checking a number of possible routes from map data alone, it will usually be possible to narrow the choice down to a few alternate possi-

bilities, thus greatly reducing the amount of field work.

If good topographical maps are available, the routes to be studied are then drawn on the maps and preliminary profiles prepared. The actual field survey then becomes a matter of traveling along each path making checks on terrain elevations at critical points. Information about the types, sizes, locations, and characteristics of any obstacles such as buildings, trees, or water towers along the path, and for a short distance on either side of it, must be gathered. Other pertinent data such as the location of large bodies of water or flat bare fields which could provide efficient reflection points should be recorded.

Path Profiles

A profile chart is prepared after tentative antenna sites have been selected, and the relative elevation of the terrain between these sites has been determined. It is not always necessary to make a complete profile of all the intervening path, in some cases only the end sites and certain hills or ridges that might obstruct the path are needed for the profile.

Microwaves generally travel in straight lines; however, the beam may be bent toward the earth by atmospheric refraction. The amount of bending, which varies with atmospheric conditions, must be considered when drawing the profile. The degree and direction of bending is called K. This factor, K, multiplied by the actual earth radius, R, gives a curve equivalent to the relative curvature of the microwave beam with respect to the curvature of the earth.

This relative curvature can be shown graphically; either as a curved earth with radius KR and a straight-line microwave beam, or as a flat earth with a microwave beam having a curvature of KR. The second method permits illustration and investigation of several K values on the same chart, and eliminates the need for special earth curvature graph paper (see Figure 3).

A path profile plotted on rectangular graph paper with no earth curvature, and with a microwave beam drawn as a straight line between the antennas represents conditions when the beam has a curvature identical to that of the earth, and the equivalent earth radius, K, is equal to infinity. This is one of the extreme conditions that must be investigated when making a study of the effect of abnormal atmospheric conditions on microwave propagation. To complete a propagation study, it is necessary to show the path of the beam relative to the earth, for other expected values of K. In all cases, it is of interest to study the path under normal atmospheric conditions when K is equal to 4/3. Standard templates are used for the different values of K so that the transmission engineer need only gather the information about the intervening terrain for each profile needed (see Figure 3).

Field Survey

Final site selections must be made after preliminary planning has determined such things as operational requirements, expansion potential, reliability requirements, and costs; the points to be served have been fixed; the most likely paths to serve these points have been narrowed by map

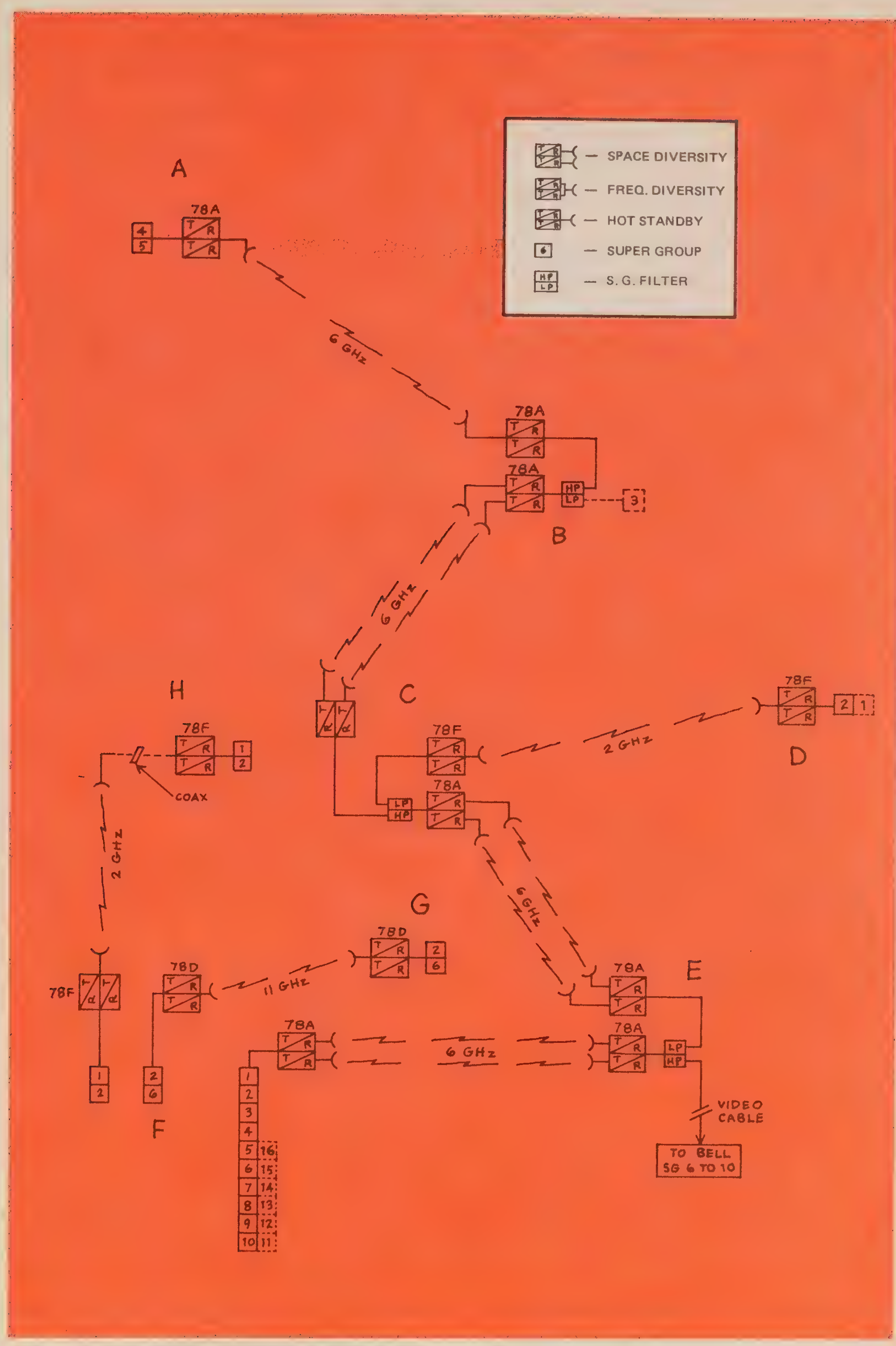


Figure 2. In laying out a microwave system, a block-and-level diagram is prepared showing all the required repeater stations.

studies; and the required system capacity has been determined. At this point the actual field survey will be undertaken.

The field survey includes more than the name implies. Actual elevation measurements, judgements of the actual terrain along each path and data concerning obstructions and possible reflections are recorded. The existence of paralleling or intersecting foreign systems and interference possibilities are indicated and the various data concerning regular and alternate sites are obtained. The preliminary profiles made from the map study become a tool for the field survey.

Probably the central problem of the microwave engineer, when he goes out in the field, is the effect of the earth itself, that is, the terrain and obstructions along the path. In fact it is this combination of terrain and atmosphere, unique to every path, which determines the propagation variations.

Accurate gathering of information about the terrain and atmosphere is extremely important when engineering a microwave system. Obtaining precise path profiles and accurate data on height and location of obstacles is the foundation of all microwave engineering, and is the principal problem facing the engineer responsible for laying out a system.

Terminal sites are often locations of existing structures or facility terminals, but intermediate sites are located with considerable emphasis on factors having to do with propagation over the intermediate paths and possible interference from sources internal or external to the system. The choice of intermediate repeater sites is greatly influenced by the nature of the terrain between sites. Preliminary map studies should have narrowed down the choices. These preliminary choices are supplemented, corrected or actually replaced as a result of the field data gathered. In the absence of actual path tests, the information brought in from the field survey constitutes almost all of the factual data about the route. From this data final judgments are made to determine the expected service performance of the system when installed.

The complexity of the required path studies varies widely depending on a great many factors. In some cases it may be possible to make a simple visual determination of path clearance by optical testing methods such as flashing, optical surveying, and "balloon" flying (see Figures 4 and 5). But, in other instances it may be necessary to conduct extensive field studies of the terrain using transit, rod, and chain, or precision altimeters to locate and describe obstacles and potential reflection points along the proposed microwave path.

Accuracy

A point that needs to be emphasized is the need for precisely locating everything at the site and along the path. As well as knowing where the obstacles are located, sites for station locations must be reported to the FCC with an accuracy of ±1 second of latitude and of longitude. For this final location determination, it is often advisable to have the precise sites determined by surveying.

Apart from the accuracy needed to meet the FCC requirements, there is a

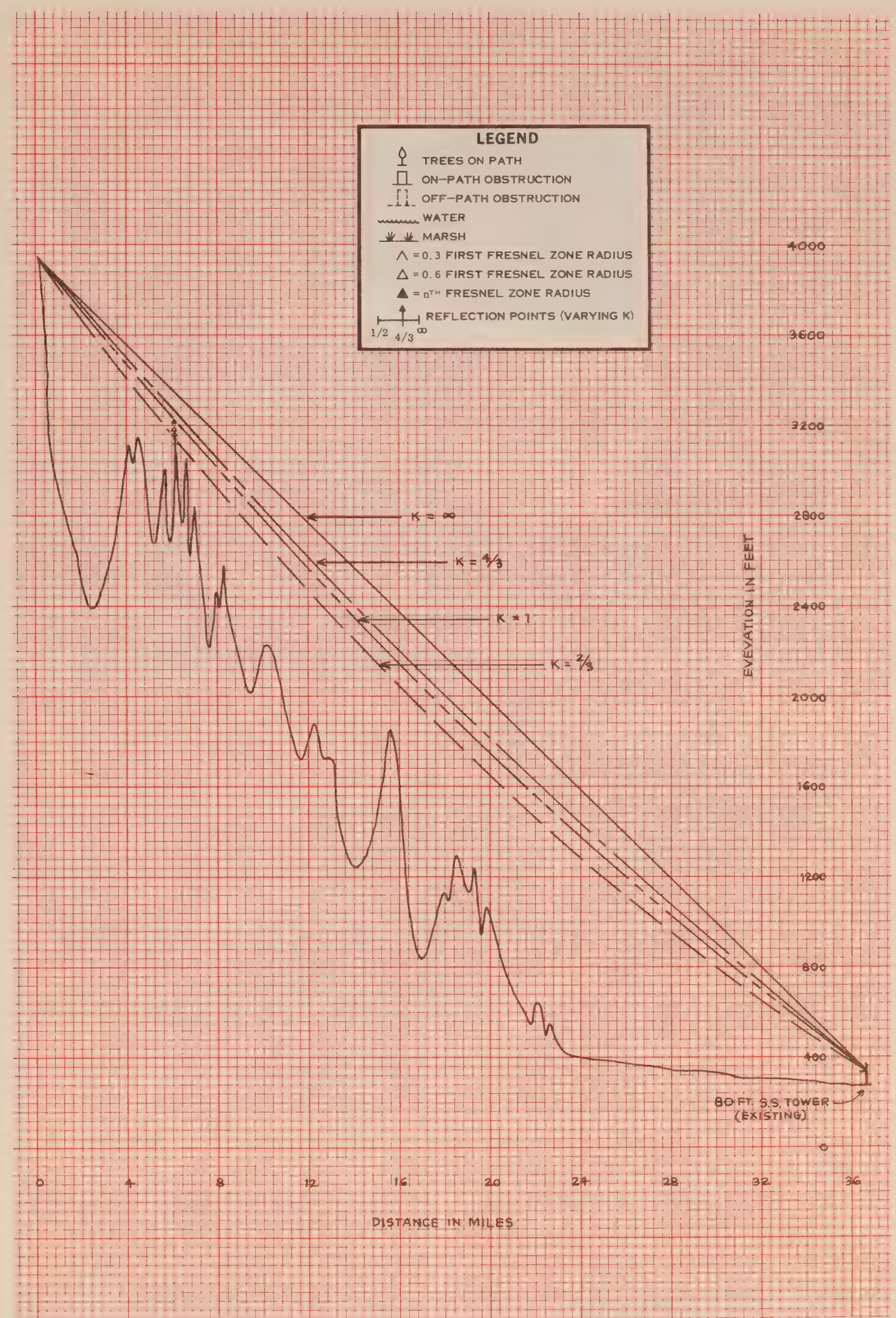


Figure 3. Plotting profile charts using a flat earth and a curved microwave beam allows the consideration of several values of K on the same chart.

Figure 4. The auxiliary equipment needed for "balloon" flying includes a helium source, VHF radio, and a calibrated, motor-driven reeling mechanism.

need for accuracy to make the overall survey data meaningful to the transmission engineer. And in this case the accuracy requirement is not only for the location of the end points, but the location of the path itself. Very often the question of whether a particular terrain feature, or a building constitutes an obstruction hinges on whether or not it is exactly on the proposed path. If the obstruction is as little as a hundred feet off the path it may not have any bearing on the propagation. In the middle of a path it is possible to

have an uncertainty of 100 feet as to where the path actually lies. One solution is to locate the path as accurately as possible, then assume that anything lying within about a quarter of a mile on either side of it is "on path." This sort of assumption is usually feasible in flat terrain, though it may sometimes make a path look bad when it really isn't. But in hilly terrain moving a path a quarter of a mile to one side may make it a totally different problem; so, the need for accuracy is great.

It is not always easy to determine the exact location of a path. The problem is accentuated in roadless country, or in hilly and forested terrain. In such cases the microwave transmission engineer often has the feeling he's working in the dark, and a great deal of ingenuity may be needed to find out when and where is he on path.

Near-end obstructions are just as important as distant ones and in cities, for example, it will often be found that moving a site only a few feet may make the difference between having a clear path and an obstructed one. Getting in and out of big cities is often one of the most complicated parts of the microwave transmission engineering. Often tower height limitations make it mandatory that the first hop be a short one.

Additional Information

In addition to the path profile and clearance data, the microwave transmission engineer must compile a great deal of other pertinent field information, some of which affects present transmission engineering decisions and some of which is for future microwave system considerations.

The location of each repeater site should be given by latitude and longitude, by word description, and by access route. The description should be adequate to allow someone who has never been there to locate not only the site, but in many cases the exact point where the tower will be located. Staking of sites is not always possible due to future negotiations regarding land purchases.

The accessibility of the site should be given by such information as the condition of existing roads, or amount of new road construction required. When construction of access roads is required, estimates of the relative difficulty or ease of construction should be indicated, since road costs can have a sizeable effect on the overall economics of the site. The type of vehicle necessary to gain access to the site should also be determined along with the travel time from the nearest highway to the antenna site.

Part of the accessibility of the site deals with the availability of commercial power. If power is not readily available at the site, details regarding the feasibility and probable cost of constructing new power lines should be determined.

FCC and FAA regulations require that microwave site applications provide such information as the distance and direction to the nearest airport, commercial or military. This information can be determined primarily from the study of maps.

Details about the site are also helpful before any construction takes place. For example, it is helpful to

Figure 5. Optical surveying using a transit is a frequently used field technique.

know the type of soil on which the building and tower will be constructed. When considering the costs of developing the site it is also helpful to know the amount of clearing and leveling, if any, that must be done to the site before it can be developed. Along these same lines it is essential to know the total amount of land available for building and tower construction, and if there are any building codes that might dictate the type of construction that has to be used. And, existing buildings already on the site

might be used for the station, thereby minimizing construction costs.

It is also necessary to have some indication of the types of weather that can be expected in the site area and along the path. Specific bits of information such as the amount of snow and ice accumulation that can be expected, the maximum expected wind velocity, and the temperature range expected are all useful.

It is highly advantageous, where possible, to establish any new stations on sites which are either on or close to existing roads and power lines. In some cases a site having such a condition may be preferable to another site which would allow lower tower heights, but would require considerable road and power construction.

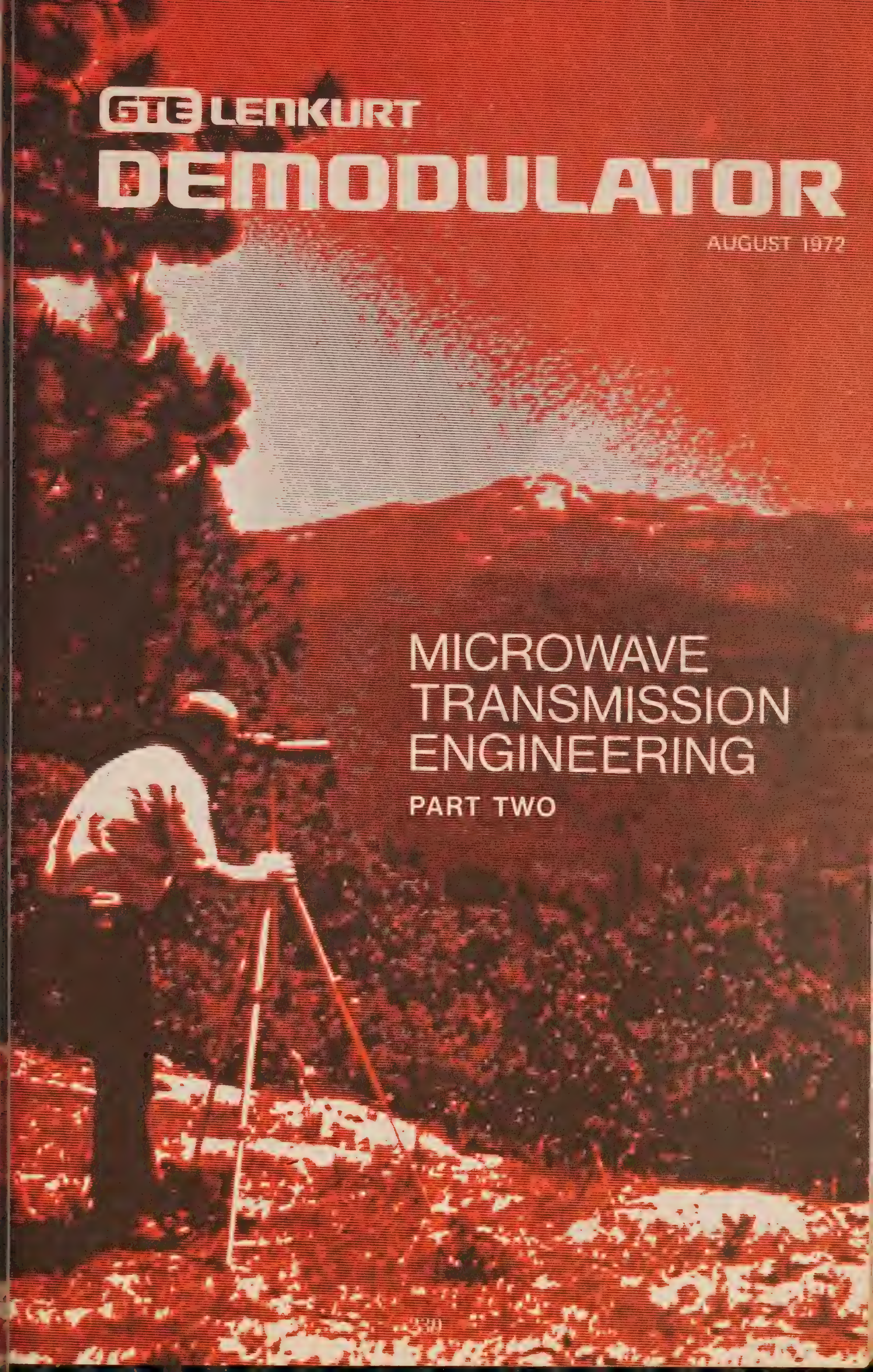
In hilly or mountainous areas, elevated sites can often be found which would be ideal repeater locations from the transmission standpoint, but so inaccessible that inordinate expense would be involved in bringing in roads and power. In more fortunate situations access may be available, or a site already developed by someone else may offer possibilities.

In evaluating access difficulties, "worst season" conditions should be used. Snow, ice, and mud can be rather impressive barriers to easy entrance and exit. Snow and ice, in large quantities, can also have harmful effects on microwave stations, towers, and antenna systems, and in areas where such things are common, careful consideration should be given to their possible effects.

A problem which must be considered in planning new microwave systems or in adding new frequencies is coordination with existing systems operating in the same microwave bands. Such coordination will often involve only suitable frequency selections for the system, but in heavily congested areas it may not be possible to obtain completely clear frequencies and in such cases it may become necessary to provide enough geographical and angular separation between the paths of the proposed system and existing systems to allow them to exist in harmony with each other.

More on Techniques

The next issue of the GTE Lenkurt Demodulator will cover the actual techniques that are used to support these principles of microwave transmission engineering. Such things as the methods used to determine path clearances, how to perform path tests, and frequency coordination studies will be covered in detail.



GTE LENKURT

DEMODULATOR

AUGUST 1972

MICROWAVE TRANSMISSION ENGINEERING

PART TWO

The basic information needed for microwave transmission engineering is essentially the same as it was twenty-five years ago, but the techniques used to gather and refine this data are constantly being improved.

Once path data has been gathered for a proposed microwave path, it is a routine engineering matter to determine antenna sizes and tower heights, to plot profiles, and to calculate system performance. Path surveys are therefore conducted for two purposes: first, to ensure that the system will operate with the necessary performance; and second, to keep the costs down to the lowest possible level and still maintain performance. Consequently, the path survey is used principally to determine minimum tower heights, minimum number of repeaters, and locations which do not result in excessive building, access, power, and maintenance costs.

Path Contouring

There are several methods used to determine elevations at the sites and along the path. Each method has merits, and each if properly used can provide the required information to make an accurate profile. Four methods are discussed, the choice is made on the basis of survey team experience, economics, and other factors.

The first method is the time-proven land survey; using transit, level, rod and chain. This method can provide all the information required for both site and path engineering, however it requires more manpower than a survey team normally used for microwave surveying, thus incurring higher costs.

A second method utilizes a combination of transit and altimeter. Surveying for sites (property lines, towers, etc.,) with transit and precision altimeter, keeps the survey team to a two-man effort.

For determining the elevation variations along a microwave path a sensitive and accurate altimeter is required, one capable of measuring to within five feet. This requires a precision instrument in good working condition. The use of an altimeter requires meticulous and careful work, and good records, plus some cooperation by nature, to get good results. Because of the potential sources of error, it is highly desirable to do a good deal of checking and cross-checking. On some days, unusually large and sudden barometric pressure changes may occur,

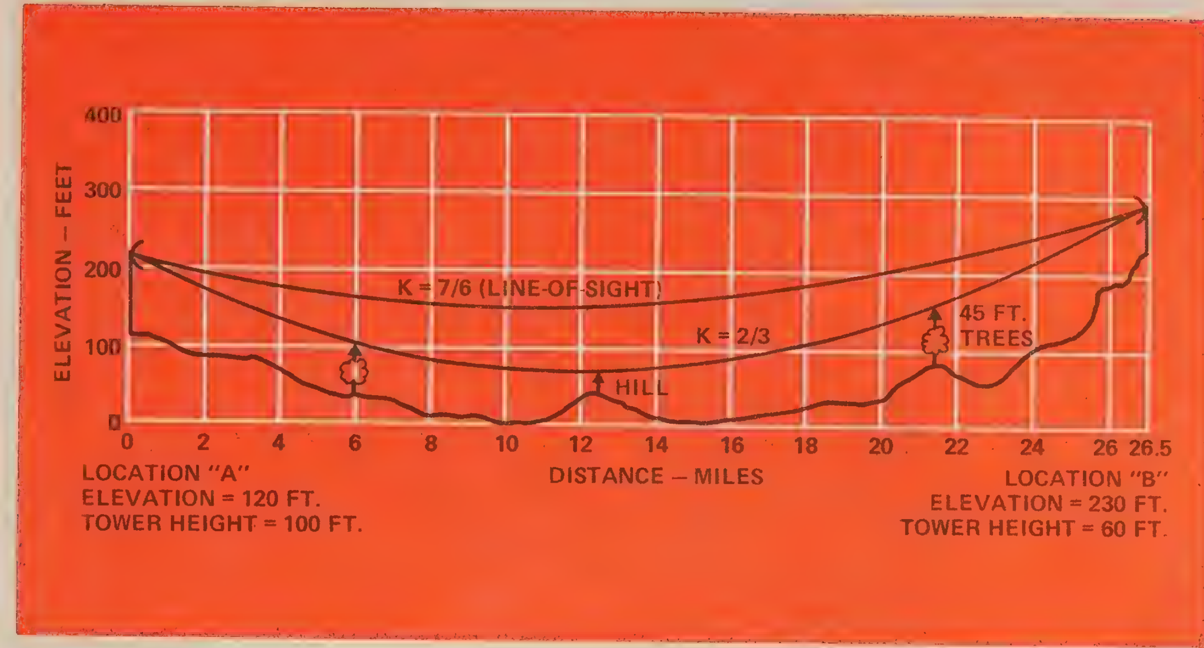


Figure 1. A tentative path profile can indicate the critical points that should be checked in the field.

and in some cases it may be necessary to discard all the results and wait for a more favorable day. Maximum use should be made of bench marks and other points of known elevation which can be obtained from such sources as the Geodetic Survey.

In well-developed areas where networks of roads exists, it is often possible to check all, or almost all of the critical points by criss-crossing back and forth along the route. Tentative path profiles often show that there are only a few points which are really critical insofar as clearance is concerned (see Figure 1). In general it is only the high points which are of interest, since they control tower height requirements.

The most accurate altimeter method for determining elevations is known as the two-altimeter method. The process involves placing both altimeters at the nearest bench mark and calibrating them exactly alike. The altimeter measures changes in atmospheric pressure so both have to be calibrated to measure from the same reference point — in this case, the pressure at a known elevation such as a bench mark.

The work should be done during stable weather conditions, and in the period from at least one hour after sunrise to one hour before sunset. One altimeter remains at the bench mark throughout the measuring period. If the stationary altimeter is manually read, readings of temperature and elevation (which would indicate pressure changes) should be taken every five minutes until the roving altimeter returns. The readings of the two instruments should then be compared. If the readings of the two instruments differ by as much as five feet after temperature stabilization, the instruments should be recalibrated and the survey repeated.

Figure 2. The barograph and the roving altimeter are calibrated to a known elevation at the beginning of the survey and rechecked at the conclusion.

If the fixed altimeter is a recording type rather than a manually read type, manpower is conserved. The principle of operation is the same as with two manually read instruments. Figure 2 shows the calibration of two altimeters.

The roving altimeter is used to measure the elevations along the path. At each measuring point, a record is made of the mileage, the temperature, the time, the trees and other obstructions, and any other terrain features of interest in preparing the profile. The roving altimeter should be used as close to the bench mark as possible. As many bench marks should be used as

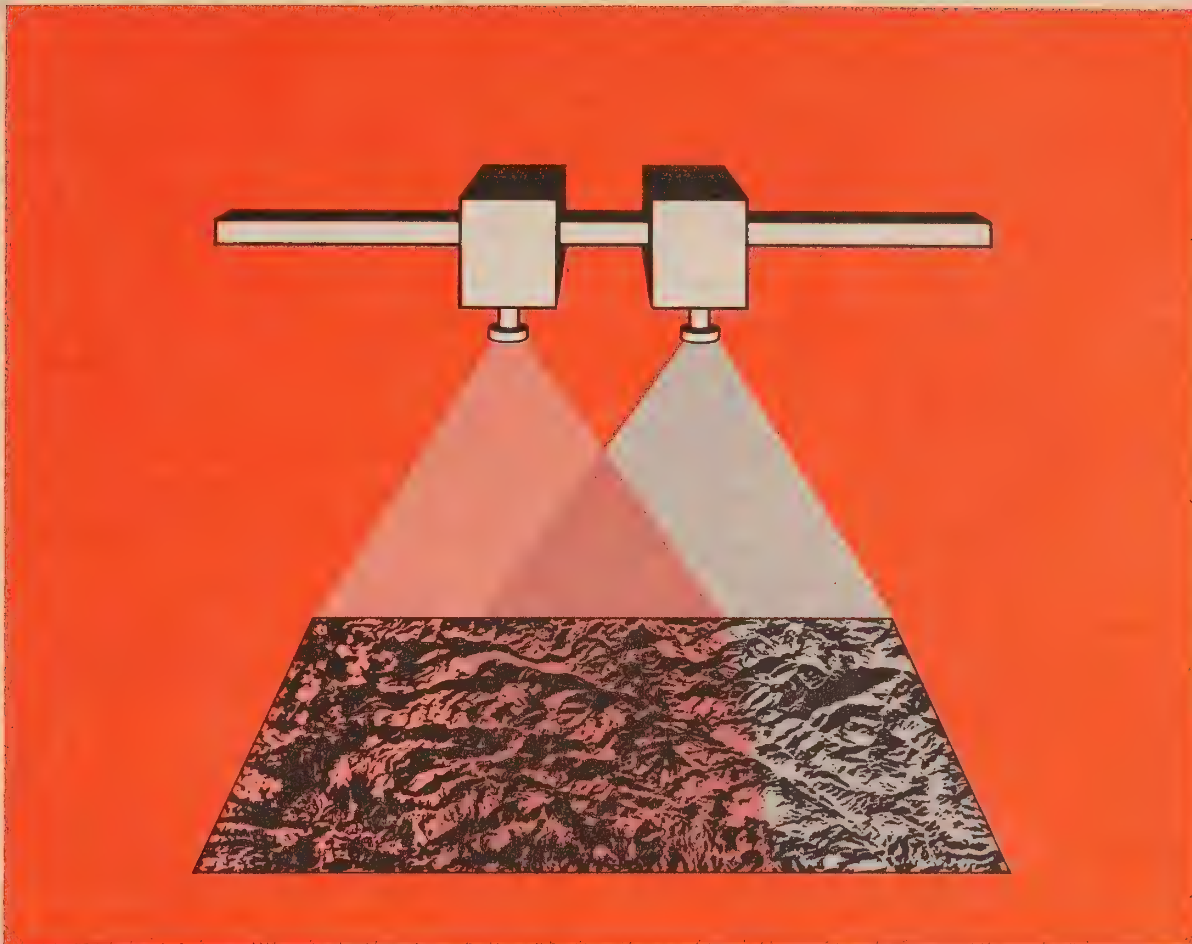


Figure 3. Stereoplotting is used to compile topographical map detail from a pair of aerial photographs that overlap by 60%. The information gathered from the photographs gives only relative elevations and must be compared to known elevations to determine absolute values.

necessary to accurately survey all the path. The final measurement at the bench mark on each measuring trip provides a reasonable check on the data. The team then moves to the next bench mark and proceeds in the same manner. If there is only one bench mark near a path, it is desirable to create secondary bench marks.

The computed elevations for the points measured are determined after corrections have been made for variations in temperature and pressure. There should be enough points measured to fully describe the profile of the path.

If the terrain along the proposed path is not conducive to the altimeter method, aerial photography provides a third method. The relative elevations are then determined by stereoplotting (see Figure 3). The accuracy of this method is good but the cost is substantially higher than the altimeter survey method. In a case involving a fairly large number of profiles, and where added manpower would be required to conduct an altimeter survey, aerial photography should be investigated as an alternative to a field survey. It should be borne in mind however, that the field altimeter team or survey

party brings in more than just elevation data, and such additional information is necessary in any event. A fourth method of profiling involves flying over the path with equipment which measures clearance using the radar principle. This is also relatively expensive and has certain limitations, such as site staking.

Obstacles

In mountainous country where all of the intermediate terrain along the path is rough and timbered, a field altimeter survey should not be attempted if adequate clearance can be determined by other means. This can often be determined by "flashing" the path. Flashing involves two parties, one at each site, preferably equipped with two-way VHF radios for communications. Lacking the radio, it is possible to coordinate by time.

The flashing is done either with a mirror reflecting the sun or with a high-power searchlight. Some engineers prefer the searchlight since it is easier to spot than mirror flashing, even in broad daylight, and the searchlight has the advantage of not being dependent upon sunshine (see Figure 4). When flashing with either a mirror or searchlight, extreme care must be taken particularly at night or in the winter to ensure that the beam is not passing through bare trees which will become obstacles when they leaf out in the spring.

For flashing to be effective, it is necessary for the sending party to establish the exact direction to the opposite site, and to concentrate the flashing along that path. One way of doing this is to drive stakes to mark the exact path and then gradually raise the beam until it is level with the distant horizon. This process should be repeated at frequent intervals until the far party is known to have seen and identified the flashes. At the far site the flashes will appear very large. A transit set up at the far site can be very useful, both for observing the flashes and obtaining accurate path sitings. The terrain clearance can be visually, but not precisely, established during the flashing. It is important that any known ridges and obstacles which appear close to the line-of-sight path be checked for elevation, and adequate clearance computed.

Another visual technique that may be utilized to check for adequate clearance in rugged terrain is balloon flying. In this process a helium filled, streamlined balloon is raised on a measured line from one site until it can be seen at the next site. In this way it is possible to determine the necessary tower heights. Since the balloon is quite stable, a reflective shield attached to the line can be used as in the flashing technique to record the sitings between antennas.

Complete profiles for paths and alternate paths should be carefully plotted with obstructions and estimated antenna height requirements, together with all the details relating to possible interference. In addition, all pertinent supporting data, altimeter readings, maps, and bench-mark information should be furnished.

The final profiles provide the basis for microwave system engineering, including the final selection of paths, antenna elevations, antenna sizes and configurations, and computation of

Figure 4. A searchlight provides an efficient method of "flashing" a path in daylight and poor visability at distances up to 30 miles.

expected signal strengths, fade margins, and system noise.

After final selection of the precise locations for the towers and antennas, the latitude and longitude of each location must be carefully determined. For systems under FCC jurisdiction, the rules require that the coordinates of the antenna or final radiating element be determined to within one second of a degree. Path azimuths should be determined to the nearest tenth of a degree of arc and path distances to the nearest tenth of a kilometer. Present GTE Lenkurt procedures involve the use of computers for these calculations.

Interference

Within the past year or two there have been several significant developments affecting the frequency coordination problem within the U.S.A. Since the FCC opened the door to the so-called specialized common carriers, there has been a large increase in the number of companies filing for microwave routes. Furthermore since many of the new systems tend to follow the same routes – often already congested with existing microwave systems – the interference problem is amplified even further. There is also a growing problem of coordination between terrestrial microwave systems and earth-

satellite systems sharing the same bands. At present there are only a handful of earth stations within the U.S.A. which provide international service via satellite; these stations are located in isolated and well shielded areas. But with the imminent establishment of domestic satellite operations there will be a large increase in the number of earth stations, and they will probably be located close to large population centers and hence in areas with large concentrations of terrestrial microwave systems.

Two relatively new calculation requirements are now being or will soon be imposed on frequency planners as a result of sharing frequencies with satellite systems. Terrestrial microwave stations operating in the 6-GHz common carrier band will in general not be authorized if the antenna beam points within ±2 degrees of any point on the geostationary satellite orbit. (Under certain conditions such a station may be authorized on a waiver basis, but with limitations on the allowable radiated power). A new coordination requirement is now being established to take into account the effects of interference via precipitation scatter, when the beams of a terrestrial station and an earth station intersect with a common volume in the lower atmosphere.

Developments such as these, together with the great number of stations which must be studied and the sheer bulk of the computations which must be made in finding interference-free frequencies for a proposed new route have resulted in a strong trend toward the use of computers to do as much as possible of the searching and calculating processes involved, leaving only the most difficult cases to be resolved by detailed manual calculations.

The requirement to establish and maintain an up-dated base of existing and applied for stations, crucial to an efficient computer operation, has some side benefits. Much information helpful toward the establishment of new routes can be gleaned from a study of existing routes in the area.

There have also been some encouraging recent developments resulting from increased cooperation and interchange of information among users, manufacturers, and the FCC about many of these problems. The common carriers, for example, have been able to reach fairly good agreement among themselves on how to define and calculate an acceptable level of interference into a given microwave system. For the industrial (private user) microwave bands, with comparable but not identical requirements, a joint task force of the E.I.A. (Electronic Industries Association) and the O.F.M.C. (Operational Fixed Microwave Council) has been successful in developing a set of acceptable interference criteria for microwave systems. A recent informal engineering meeting held under FCC sponsorship developed essential agreement on methods to be used in calculating whether or not a microwave beam intersects the geostationary orbit. Thus, though there remain many areas which need further effort, a useful pattern seems to be emerging, looking toward closer co-operation among user groups, manufacturer groups, and the regulatory bodies in finding solutions to the ever increasing interference problems.

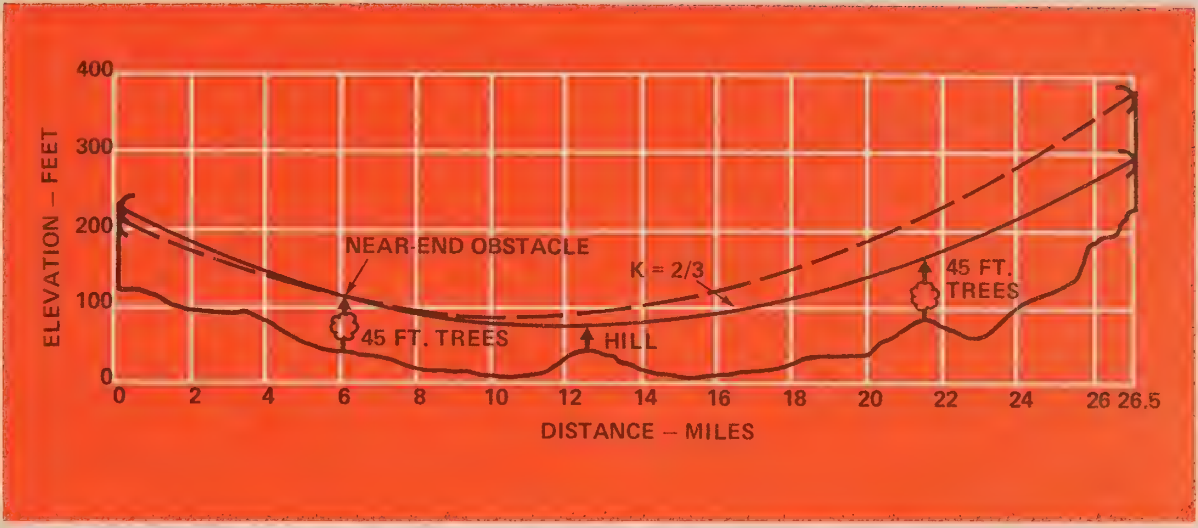


Figure 5. An obstacle near one end of the path will mean that changing the antenna height near the obstacle will result in a larger change at the other end of the path.

An unfortunate but perhaps inevitable consequence of the increasingly complicated calculations required is that it is becoming more difficult for a small user to do his own frequency planning, even if he has a knowledgeable engineering staff. Such users are being forced to turn either to the microwave manufacturers or to companies specializing in providing frequency analysis.

Profiles

After a path has been selected and surveyed, and all the path data collected, the next step is to plot the path profile and determine the required tower heights. The elevations of the end points, and of any high points or controlling obstructions should be carefully plotted. Lower intermediate points can be simply sketched in, if they will not effect clearance problems. Trees, buildings, or other obstacles which have been identified along the path should be accurately plotted.

It is always possible in theory to reduce the required height at one end by raising the height at the other, but whether it will be advisable to do so will depend on where the critical points are located. Often there will be only a single critical point near the center of the path. In such a case the trade-off, at least over a certain range, will be on a one-for-one basis. In other words, raising one tower by X feet allows the other to be lowered by X feet.

If the controlling obstruction is near one end of the path, the trade-off is no longer one-for-one. The path can be thought of as a lever, or an unbalanced seesaw, pivoting around the obstruction point (see Figure 5). Thus a change in height at the end nearer the obstruction can only be counterbalanced by a larger change at the other end, the relation between the two depending on the relative distances.

If there are several controlling obstacles scattered more or less along the path, there would be almost no trade-

GTE LENKURT INCORPORATED

CUSTOMER ____________________

ENGINEER ____________________ SYSTEM ____________________

DATE ____________________ PROJECT NO. __________ FREQUENCY __________ GHz

1	SITE	
2	LATITUDE	
3	LONGITUDE	
4	SITE ELEVATION, AMSL	Ft
5	TOWER HEIGHT (Nom.)	Ft
6	TOWER TYPE	
7	AZIMUTH FROM TRUE NORTH (G.C.)	
8	PATH LENGTH (G.C.)	Miles
9	PATH ATTENUATION	dB
10	RIGID TRANS. LINE	Ft.
11	FLEXIBLE TRANS. LINE	Ft.
12	TRANSMISSION LINE LOSS	dB
13	CONNECTOR LOSS	dB
14	R-F COMPONENT LOSS	dB
15	RADOME LOSS, TYPE	dB
16	TX LINE + ANT RETURN (VSWR)	dBRL
17	TOTAL FIXED LOSSES	dB
18	TOTAL LOSSES	dB
19	PARABOLA HEIGHT, ℄ AGL	Ft.
20	PARABOLA DIAMETER	Ft.
21	REFLECTOR HEIGHT, ℄ AGL	Ft.
22	REFLECTOR SIZE, TYPE	Ft.
23	PARABOLA-REFL. SEP.	Ft.
24		
25	ANTENNA SYSTEM GAIN	dB
26	TOTAL GAINS	dB
27	NET PATH LOSS	dB
28	MIN. TRANSMITTER POWER	dBm
29	MED. REC. POWER (±2 dB)	dBm
30	NOISE THRESHOLD (KTB +NF)	dBm
31	THEORETICAL RF C/N RATIO	dB
32	REC. MUTE POINT	dBm
33	FADE MARGIN (To Mute Point)	dB
34	RELIABILITY,	
35	PROFILE NUMBER, K, F.	
36	UNFADED NOISE	pWp0

6 SS Self Supporting, G – Guyed, S – Stub, R – Roof Mount, W – Wall Mount, E – Existing
7,8 Great Circle Computations Based on Coordinates Shown (Inverse Position)
10,11 R – Rectangular, E – Elliptical, C – Circular, X – Coax
14 C – Circulator, H – Hybrid (Split), F – Filter, A – Attenuator, S – Switch, X – Coax
15 U – Unheated, H – Heated, F – Feed, P – Planar
20 X – Cross Polarized, H – Shrouded, D – Dual Frequency, G – Grid
22 C – Curved, F – Flat, S – Special Tilt Angle, R – Rectangular, E – Elliptical
23 For Billboard Passive, Show the 2 α Angle Here
34 Diversity: F – Frequency, S – Space, HY – Hybrid, HS – Hot Standby, N – None
Reliability: Weighted Normal Propagation Reliability

EQUIPMENT

TYPE __________ , __________ MHz I.F.
__________ EMPHASIS __________ KHz SLOT
__________ CHANNELS __________ KHz DEV/CH
± __________ dBm 0 LOADING __________ dBRL

B–758 (5–72)

Page___ of___ Issue___

Figure 6. Microwave path data calculation sheets are used to determine both qualitative and quantitative design objectives and once completed they become the system documentation for equipment specification and installation.

offs available, since the heights shown at each end would be required to clear the near-end obstructions, even if those at midpath did not exist.

Finding the best combination of antenna heights and assuming that the clearance is adequate, essentially completes the path solution in cases where there is no likelihood of reflections. When surfaces of high reflectivity such as bodies of water, salt flats, or very smooth terrain exist along the path, some additional work may be needed.

Reflections

The "elegant" way to check for reflections is to run height-gain tests to determine the extent of the reflection problem, and the best antenna heights. This is, however, a very expensive process, particularly if the necessary portable towers and test equipment are not readily available and must be purchased or leased.

There are other things short of path testing, which may be done to protect against reflection problems. The first thing to be done, after choosing tentative antenna heights purely on a clearance basis, is to check whether this set of heights has a clear, unblocked reflected path from the potential reflective surface, under expected operating conditions. There may be some terrain feature which blocks one of the reflected paths, but allows adequate clearance for the direct path. This would be a desirable situation which would effectively remove the reflection problem.

If it is found that the initial antenna choices do allow unblocked reflections to exist, it may be possible to change antenna heights, or in some cases even the station locations.

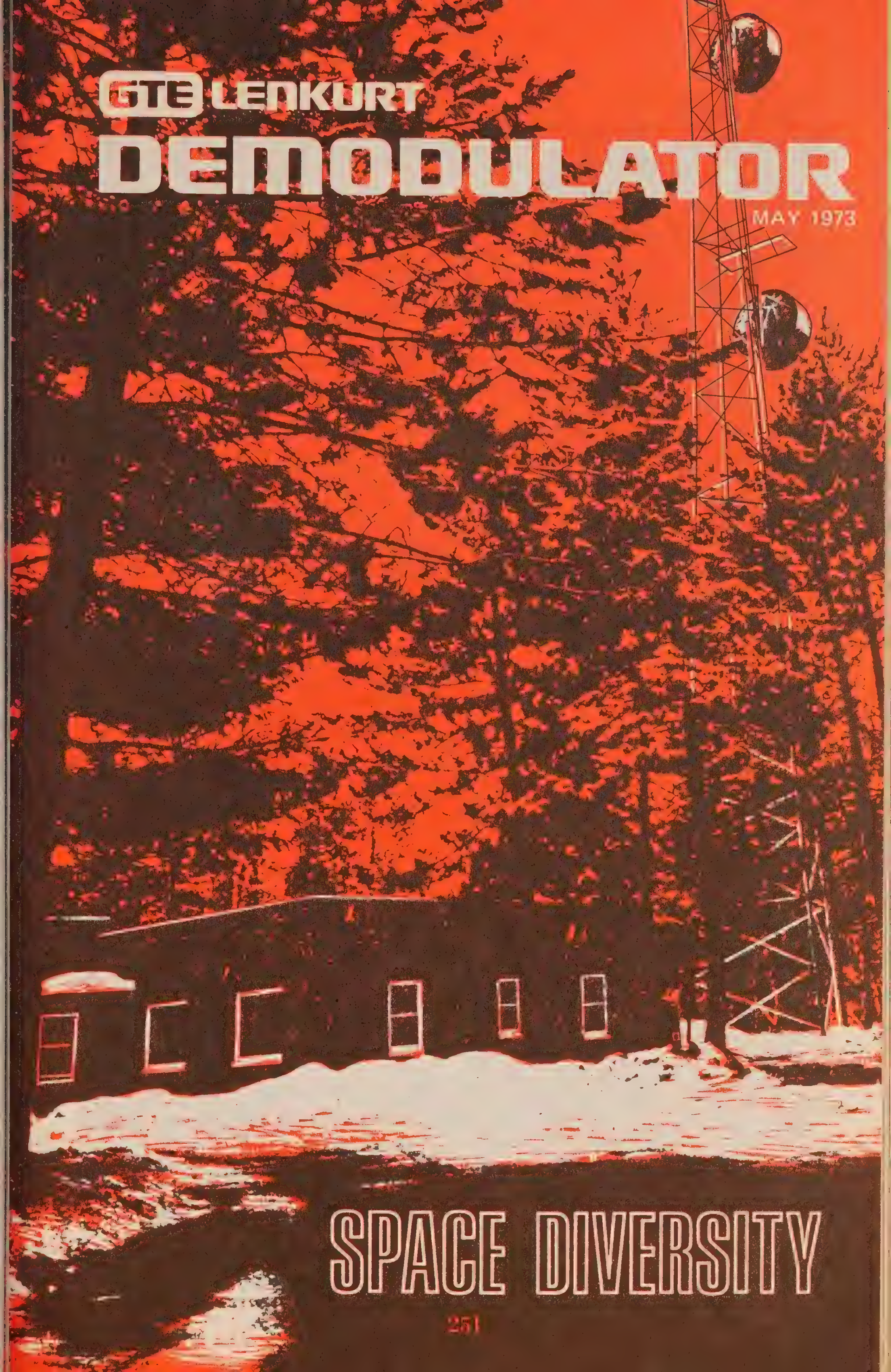
In some cases, it may not be possible to avoid having a potentially reflecting path. In such cases it may be necessary to resort to space diversity to alleviate the problem.

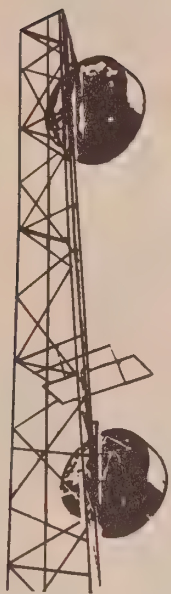
One of the last things to be done is to enter the appropriate data on a microwave path data calculation sheet (see Figure 6), together with the pertinent equipment and performance parameters, and calculations made to determine the amount of antenna gain needed to meet the desired unfaded signal level.

After determining the antenna configuration, the appropriate gains can be entered in the sheet, and the remaining calculations made to determine normal signal level and fade margin. On a single channel or frequency diversity path, the process is quite straight forward and self-explanatory. Some complications arise with space diversity or crossband diversity, in that there are separate waveguide runs and in some cases even separate antenna systems requiring separate calculations.

Coordinated Effort

After all the engineering calculations have been completed, the appropriate equipment to meet the desired and necessary engineering specifications must be furnished and installed. The whole operation from initial request to operating system requires a coordinated effort to result in the best possible microwave system.


GTE LENKURT
DEMODULATOR
MAY 1973
SPACE DIVERSITY
251



Space diversity has been used for over ten years by industrial microwave users, and as a result of recent FCC rulings, space diversity will be finding increased application by the common carriers.

Point-to-point microwave paths, except when quite short or when located in extremely favorable areas, are subject to fading, or fluctuations in the intensity of a received signal caused by changes in the characteristics of the propagation path or transmission medium. Appropriate measures must be taken to sufficiently minimize the effects of fading and to provide the required system reliability.

Fading can be minimized by using diversity techniques where two paths are available for transmitting the same information. These paths are chosen such that simultaneous fading is unlikely. There are two distinct diversity techniques for point-to-point, line-of-sight microwave systems – frequency diversity and space diversity. Frequency diversity uses two different frequencies to transmit the same information. With space diversity, the same frequency is used, but two antennas, separated vertically, transmit or receive the information over two different paths through space (see Figure 1).

Compared to space diversity, frequency diversity is less expensive, uses simpler equipment arrangements, and has some operational and maintenance advantages. But, new FCC (Federal Communications Commission) rules prohibit frequency diversity for common carriers unless sufficient evidence can be shown that frequency diversity is the only way to obtain the required system reliability. This new ruling was established to preserve microwave frequencies for working radio channels, since there is a high demand for microwave frequencies and only a limited supply.

The use of frequency diversity is now limited by the FCC to only one protection channel for the 4-GHz band and one for the 6-GHz band, and frequency diversity will not be authorized unless the user applying for a license has at least three working channels. An exception to the three working channels provision can be made if the user can show that a total of three working channels will be required within three years.

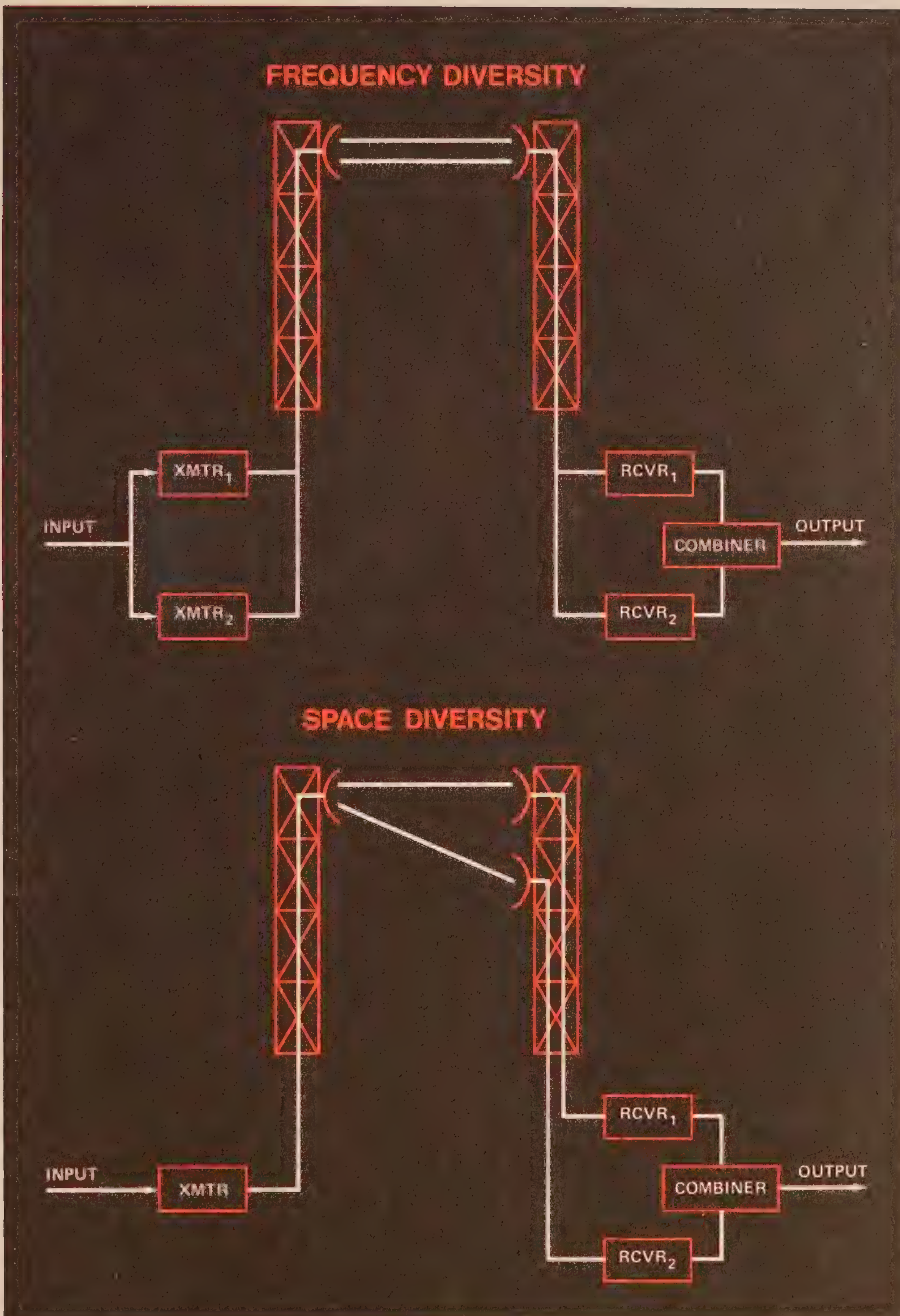


Figure 1. Frequency diversity and space diversity accomplish similar ends but they use different equipment configurations.

In effect, the FCC rules ban any form of frequency diversity for new systems in the 4- and 6-GHz bands except for routes which have at least three parallel working channels. So, except where waivers can be obtained, frequency diversity cannot be used for hops carrying only one or two working channels, and the only good alternative is to use space diversity. But, because of unfamiliarity with it, and to some extent because of some past misconceptions about it, many engineers are still skeptical about the capabilities of space diversity.

Fortunately, successful applications of space diversity do exist. Since frequency diversity has never been allowed in the industrial bands, these private microwave users turned to space diversity more than ten years ago to give them the degree of reliability needed. There are now hundreds of microwave paths operating with space diversity, mostly in the 6.575- to 6.875-MHz industrial band.

The initial decision by industrial microwave users to employ space diversity was accepted with some misgivings, since literature at that time was somewhat dubious about the protection provided by space diversity. But once the decisions were made, and the first space-diversity systems implemented, all doubts were quickly dispelled. The results obtained using vertical space diversity have generally been outstanding and have offered a superb degree of protection against fading.

The kind of diversity spacing that has been used in line-of-sight microwave systems should really be called height diversity, or vertical space diversity, since the spacing used is invariably vertical. It is generally accepted that horizontal spacings would have to be much greater than vertical spacings to provide equivalent diversity action. Although there is no direct proof, neither are there any positive tests to determine the relative effectiveness of horizontal spacings.

Published material seems to indicate that vertical space diversity is at least as good as frequency diversity, and in many cases considerably better when comparing propagation protection. Bell Laboratories' work shows space diversity substantially better than frequency diversity, particularly at 6 GHz and higher. At 4 GHz the difference is less significant, but the Bell Labs' data indicates that space diversity is still favorable. Space diversity appears to be effective in all microwave bands, even down to 2 GHz. Published Japanese experiments indicate that space diversity has an even greater protection advantage over frequency diversity than the Bell Labs' work indicates. But, because of the economic and operational advantages of frequency diversity, space diversity has only been used in a few isolated multiline cases where frequency diversity was not adequate.

Antenna Spacings

In the 6-GHz band, a non-critical vertical spacing on the order of 30 to 40 feet will provide more than adequate, indeed extremely good, diversity. In the past, controversy has developed over the proper way to select spacings. Advocates and users of space

diversity developed theories and methods for calculating "optimum" spacings, based on the assumption that the only significant contributor to multipath fading is a single discrete reflection from a path point determined by calculation from path parameters. Another view was that most overland microwave paths are subject to many atmospheric and geographic factors that cause fading, so there probably is not a way to calculate an optimum spacing.

Most users today share the view that the spacings are non-critical and need not be calculated, at least on conventional overland paths. Diversity action tends to improve as spacings increase, so from this point of view, the larger the better. However, increasing the spacing may mean increasing tower heights and costs, and the 30- to 40-foot range seems to be a good compromise in the 6-GHz band.

For many years the rule of thumb was that 40-foot spacings on a 6-GHz path with a 40-dB fade margin would give a diversity improvement, or reliability improvement, against multipath fading, of at least 100 to 1. Based on present experience this was an extremely conservative rule of thumb. Published experimental and theoretical data from Bell Labs on various aspects of microwave propagation, including both frequency and space diversity effects, indicates something like a 250 to 1 improvement for a 6-GHz path with a 40-dB fade margin. Published Japanese data indicates an even greater improvement – on the order of 5,000 to 1. Calculation methods based on Bell's work can therefore be used with considerable confidence that the results will be at least realistic, if not conservative.

Methods are now available to calculate expected, non-diversity outages, as a function of path and climate parameters, and also as a function of the improvements to be obtained by diversity. A relaxed view has been taken on antenna spacings – generally, 30 to 40 feet at 6 GHz, 45 to 50 at 4 GHz, and 60 to 80 at 2 GHz will provide enough diversity to essentially eliminate multipath fading. At 11 GHz, spacings of 25 to 30 feet would be adequate. Occasionally, on a path where water reflection or severe ground reflection is likely to predominate, a discrete calculation method can be used, but this is rare. Experience indicates that even if small spacings (15 to 20 feet) are used, there is still good diversity action, at least in the 6-GHz bands. Even 10-foot spacings have been found to provide significant reliability improvements.

Towers and Antennas

An ultraconservative method of determining tower heights for space diversity would be to increase each non-diversity height by the required amount of diversity spacing. In effect, this would mean applying the path clearance criteria for the bottom antenna to the bottom transmission path. A more reasonable approach, though still quite conservative, would be to increase tower heights by half the diversity spacing. This is equivalent to applying the clearance criteria to the path from the top antenna to the bottom antenna. The most commonly used approach however, is to apply the

basic path criteria to the top-to-top path, and a slightly more relaxed criterion to the top-to-bottom paths. Figure 2 shows three common antenna arrangements that result in three different tower heights. It is necessary under any conditions to check for problems due to near-in obstacles close to one end of the path which might block the lower antenna.

Disadvantages

There are several disadvantages to space diversity. One of these, at least in systems involving message circuits, is the impossibility of performing end-to-end testing without taking the system out of service, and is perhaps the major disadvantage of using space diversity compared to frequency diversity. Greater cost, mainly because of the extra antennas and waveguide, plus the effect on tower loading and sometimes height, is an obvious disadvantage. Space-diversity systems also require switching and sensing equipment that adds to the system's complexity.

Space diversity can be applied only on a per-hop basis, and each RF channel must have its own individual space diversity protection. This is not critical with industrial applications, where systems typically involve only a single working channel over a microwave route. But it is important to common carriers, where systems with multiple RF channels are quite common.

The "second channel" in a space diversity system is not separable, so it cannot be used on an occasional basis for other services, as is often done with a frequency-diversity protection channel. Two hot-standby, space-diversity channels are not easily converted into a 1-for-3 multiline diversity arrangement, which might occur under the present FCC rules.

Redundancy

Frequency-diversity systems have the feature that a complete end-to-end protection channel can be automatically switched-in to replace a failed channel. Regardless of whether the failure was caused by fading or equipment trouble, 100% equipment redundancy is automatically available with frequency diversity. In 1-for-N or 2-for-N multiline protection systems, switching is not required on every hop, but can be implemented over a switching section comprising a number of hops in tandem. Neither of these things is true about space diversity.

Space diversity, depending on how it is applied, may provide partial equipment redundancy or no equipment redundancy. As commonly used in industrial space diversity applications, two complete receivers are normally provided at the receive end of each microwave hop, one connected to the upper antenna and one to the lower antenna. The output of each receiver contains the full information band and in the absence of fading, loss of one receiver does not affect the output to the load, since an automatic switch or combiner is provided to leave both signals on line or to select the good receiver when fading or equipment failure causes the other receiver to lose its signal output.

Figure 3 shows a single, one-way RF channel, transmitting from A-to-B

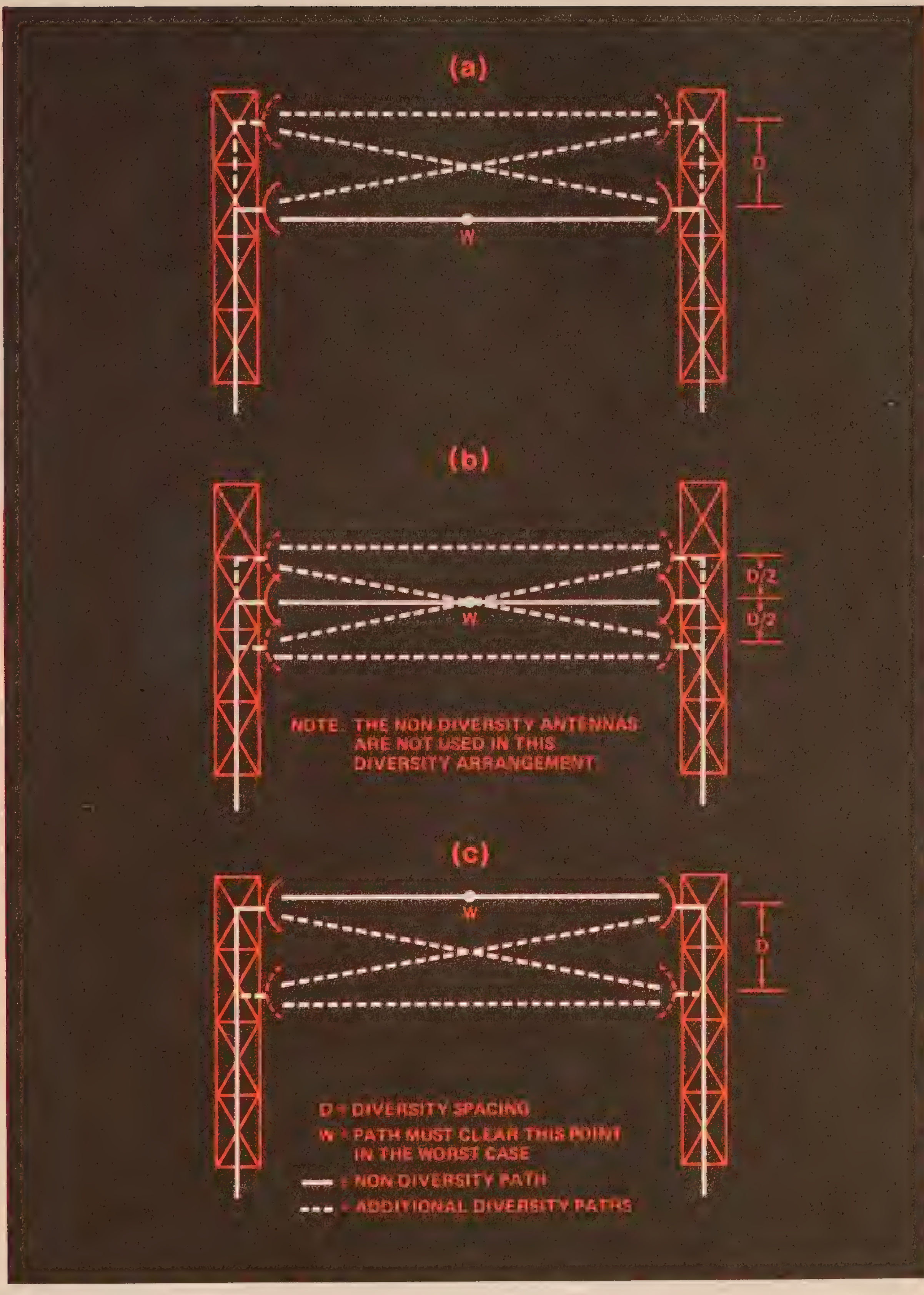


Figure 2. In the most conservative space diversity arrangement (a) with all the spacing above the existing antenna, all paths give adequate clearance; when the spacing is split above and below the existing antenna (b), three out of four paths are good; and when all the spacing is below (c), only the top-to-top path is good.

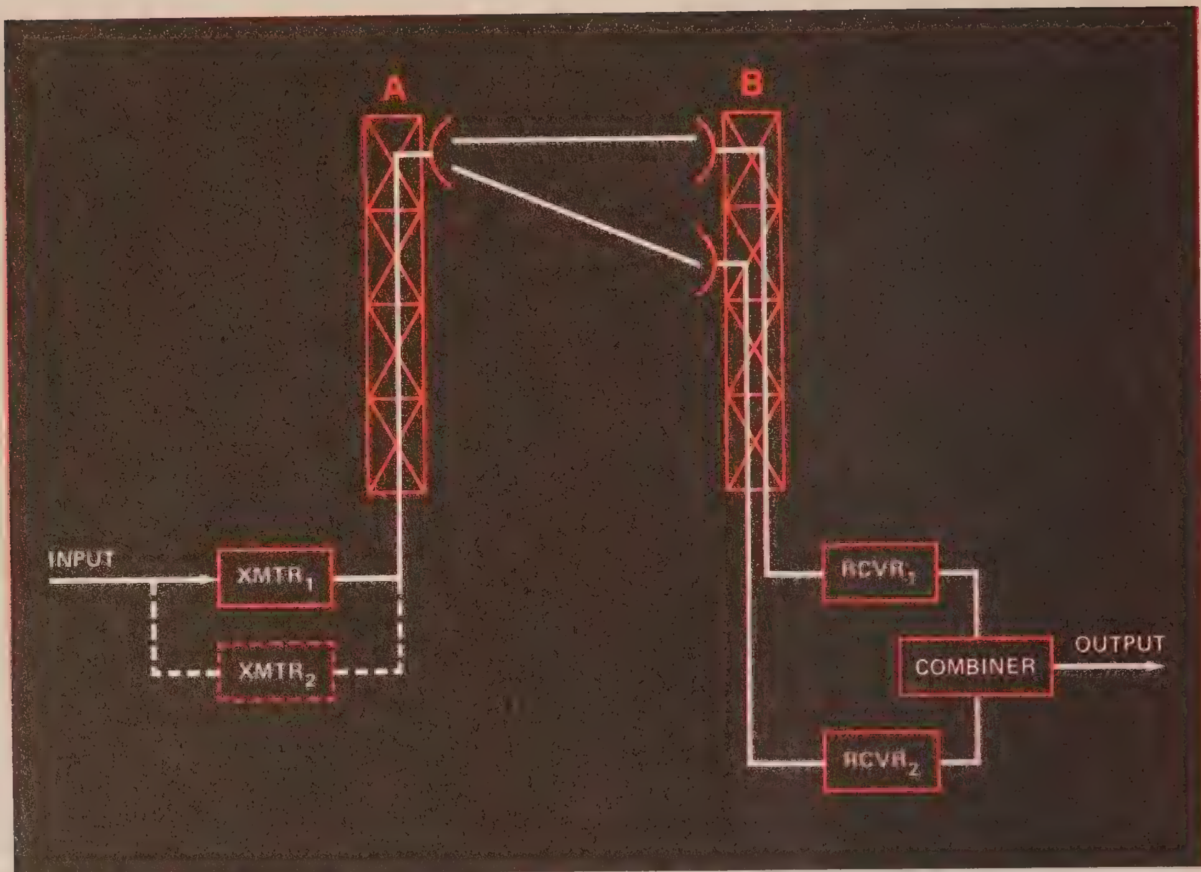


Figure 3. A one-way space-diversity systems uses diversity at the receiver end.

on a single frequency. Most systems will be duplex, and the opposite direction of transmission would require a reversed duplicate of the arrangement shown in Figure 3, transmitting from B-to-A on a separate single frequency. The space diversity in the direction of B-to-A will require two spaced antennas at the A end, in the normal configuration. In this configuration, full receiver equipment redundancy is provided, but the transmitter will not have equipment redundancy per se, since normal operation involves only one transmitter.

To provide transmitter equipment redundancy, the normal practice is to equip a complete spare transmitter, in a hot-standby arrangement, as shown in Figure 3 by the dashed lines at A. The two transmitters are identical, and typically both are modulated by the input signal and both are generating output power. But, in general, only one is "on line" at a given time. Fast-acting waveguide switches are used to keep one transmitter "on line" and the other "off line" at a given time.

When the system is of the remodulating (baseband) type, the outputs at baseband can be combined or selectively switched in such a way that a completely hitless transfer operation takes place. It is desirable to equalize the electrical lengths of the two paths to within ±5 nanoseconds to insure that the information arriving at the

combiner is in precise phase and time synchronization. Since a differential of 5 nanoseconds represents only a small part of even a high-speed digital bit, a change from one receiver to another will not cause an information change due to displacement. Equalization of path lengths is a simple process at baseband and need be done only at the time of initial lineup.

If the receivers are of the heterodyne type, with an output of 70 MHz instead of baseband, things are different. Unless synchronization techniques are used, the 70-MHz outputs of the two receivers generally will not be in phase, and may even be of opposite phase. This dictates the use of selective switching instead of combining. Near-hitless transfer is assured by using solid-state switching devices with extremely fast transfer times. A space-diversity system switched at IF frequency provides equipment protection and redundancy for the IF and RF portions of the receiver, but separate protection is needed for any demodulator and baseband equipment following the switch.

Combining and Switching

RF combining and switching have been used for problem paths on multiline systems where the frequency diversity protection has been inadequate and additional propagation protection in the form of space diversity is needed over a particular path. One way of doing this is to combine the RF signals from the two vertically spaced antennas in a "magic Tee" before applying them to the input of the receiver (see Figure 4). This scheme requires careful control of the relative positioning of the two antennas and very precise equalization of the electrical length of the two waveguides connecting the two antennas to the "magic Tee", since for proper operation the two signals have to be closely "in phase" at the microwave frequency.

At 6 GHz, a path length difference of about an inch would put the two signals out of phase by 180°, and could conceivably cause a complete loss of received signal. For this reason, systems using this method generally introduce some method of automatically sensing the degree of phase difference, and automatically controlling a variable phase shifter in one receiver branch so as to bring the phases into close alignment, even if path variations cause a considerable differential in electrical lengths. The arrangement shown in Figure 4 is a complex and expensive system to implement, and its very complexity may introduce more difficulties than those it is designed to prevent.

A considerably simpler arrangement, using RF switching instead of combining has been used to provide space diversity protection on paths with excessive fading (see Figure 5). In this arrangement, the waveguide from the two spaced antennas is fed into ports of a latching circulator which connects one of these inputs to the receiver and the other to a termination. In normal operation the receiver is connected to the main antenna. When there is a loss of signal at the receiver, as shown by the AGC (automatic gain control) voltage dropping

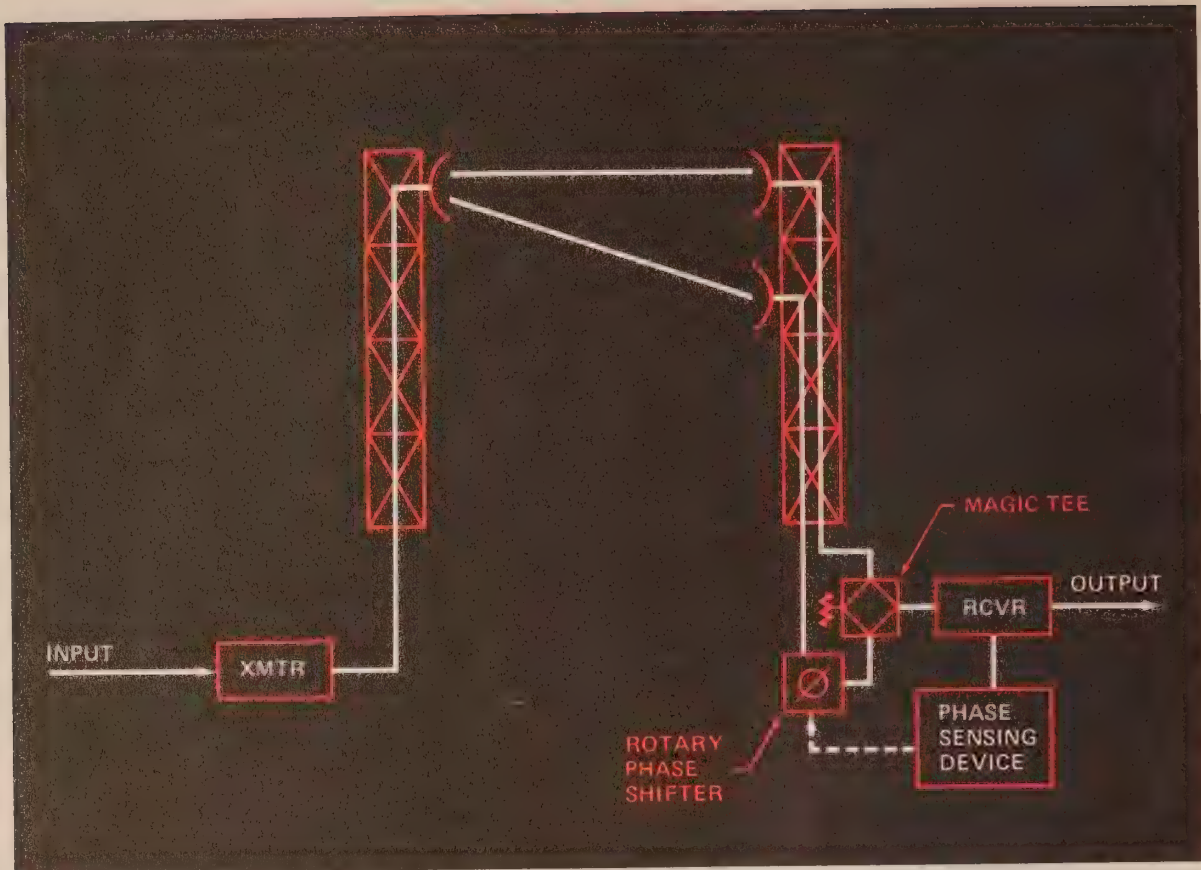


Figure 4. Space diversity without equipment redundancy uses a "magic Tee" for combining and a rotary phase shifter for automatic phase control.

to a preset level, logic circuitry switches the circulator to the other position, connecting the auxiliary antenna to the receiver.

The switch is "blind" in that there is no guarantee that a satisfactory signal will exist on the auxiliary antenna. Hence, other logic circuitry is used to tell the receiver what to do if no signal or an inadequate signal appears on the auxiliary antenna. Although some problems exist with a blind switching arrangement, it has been found to provide a substantial degree of improvement and a consequent reduction in outage time on some paths. Blind switching is also simple and easy to implement. In this arrangement, each RF channel receiver has its own individual switching and control mechanism.

There are many systems where spaced antennas are only used at one end of the path. This is particularly true when adding space diversity to existing systems whose towers have not been designed to handle extra antenna loads and in situations where one end of the path, perhaps in a built-up urban area, does not permit enough height for a second vertically spaced antenna.

To have conventional space diversity in one direction, the spaced antennas are associated with the receivers. To achieve space diversity in the oppo-

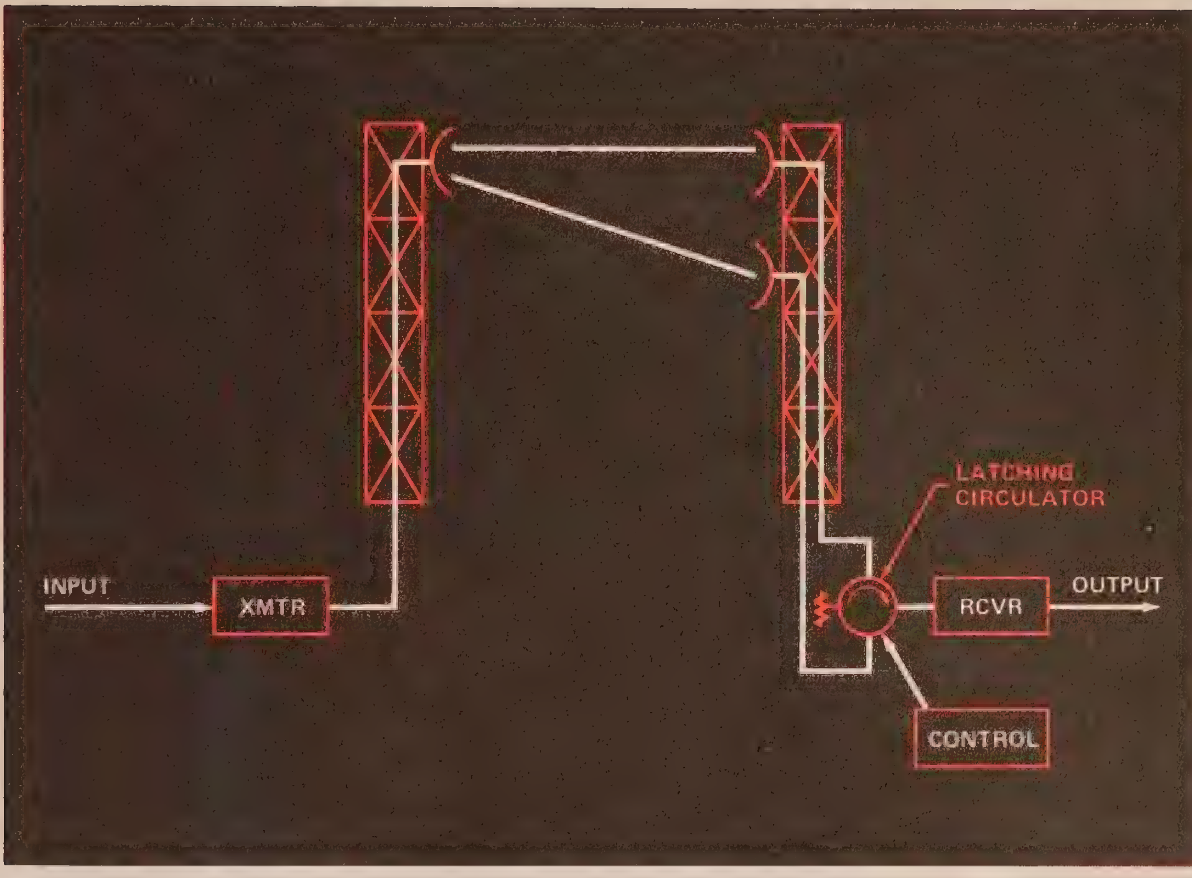


Figure 5. Space diversity without equipment redundancy can also be done using a latching circulator for "blind" RF switching at the receiver.

site direction, the spaced antennas must be associated with the transmitters.

General Use

Vertical space diversity, using relatively modest spacings, can provide extremely effective propagation protection against selective (multipath) fading. In addition to its added cost, space diversity has some operational and maintenance disadvantages when compared to frequency diversity, and these should be recognized and understood by those considering its use. Space diversity can be applied in a number of different ways, and using the experience of the industrial microwave users, several ingenious applications have already been worked out for common carriers. Other common carrier applications no doubt will develop as space diversity comes into more general use.

GTE LENKURT

DEMODULATOR

MARCH 1975

light-route radio systems

Light-route radio systems are a specialized form of communications which enable industrial and common carrier users to cut costs and still retain high quality of transmission.

The 2-GHz band is ideal for light-route communications because in this frequency range, atmospheric fading is minimal, and long transmission paths are practical. Before the advent of light-route systems, a 300-channel radio might have had to be used to carry the small number of channels often required by an industrial user, thus increasing the per-channel cost of the microwave radio system and wasting rf spectrum.

Some of the users of narrow-band, operational fixed, light-route systems include petroleum companies, railroads, and independent telephone companies. An "operational fixed" station is an FCC (Federal Communications Commission) designation used to describe a radio station at a permanent location which is not open to public correspondence, but operated by and for the sole use of those agencies operating their own radio communication facilities in the Public Safety, Industrial, Land Transportation, and Aviation services. The GTE Lenkurt 70F1 radio, for example, has been designed for use in this category of service. It is a 36-channel system which operates in the 2.11 – 2.20 GHz domestic industrial and common carrier frequency allocations. In geographical areas under the territorial jurisdiction of the FCC, this radio has been type-accepted for the types of services shown in Figure 1.

Light-Route Uses

A petroleum company can use a light-route system to control the flow of oil in a pipeline between drilling sites and refineries. The majority of the channels would probably be used for administrative functions, but the main function of the system would be in those channels used for remote and supervisory control, telemetering (the transmission of measurement data over a distance), and on-line control and switching of pipelines or pumping facilities. A typical light-route system might appear as shown in Figure 2.

Complex pipeline control networks are used in the petroleum industry to control both production and petroleum flow. Drilling and pumping operations can be remotely controlled through the use of microwave systems. The results of remote chemical analysis of production samples is used to control drilling. When the percentage

FCC RULE PART	TYPE OF SERVICE
21	DOMESTIC PUBLIC FOR COMMON-CARRIER USE
87	OPERATIONAL FIXED FOR AVIATION SERVICES
89	OPERATIONAL FIXED FOR PUBLIC SAFETY RADIO SERVICES
91	OPERATIONAL FIXED FOR INDUSTRIAL RADIO SERVICES
93	OPERATIONAL FIXED FOR LAND TRANSPORTATION RADIO SERVICES

Figure 1. Example of FCC radio type acceptance for specified services.

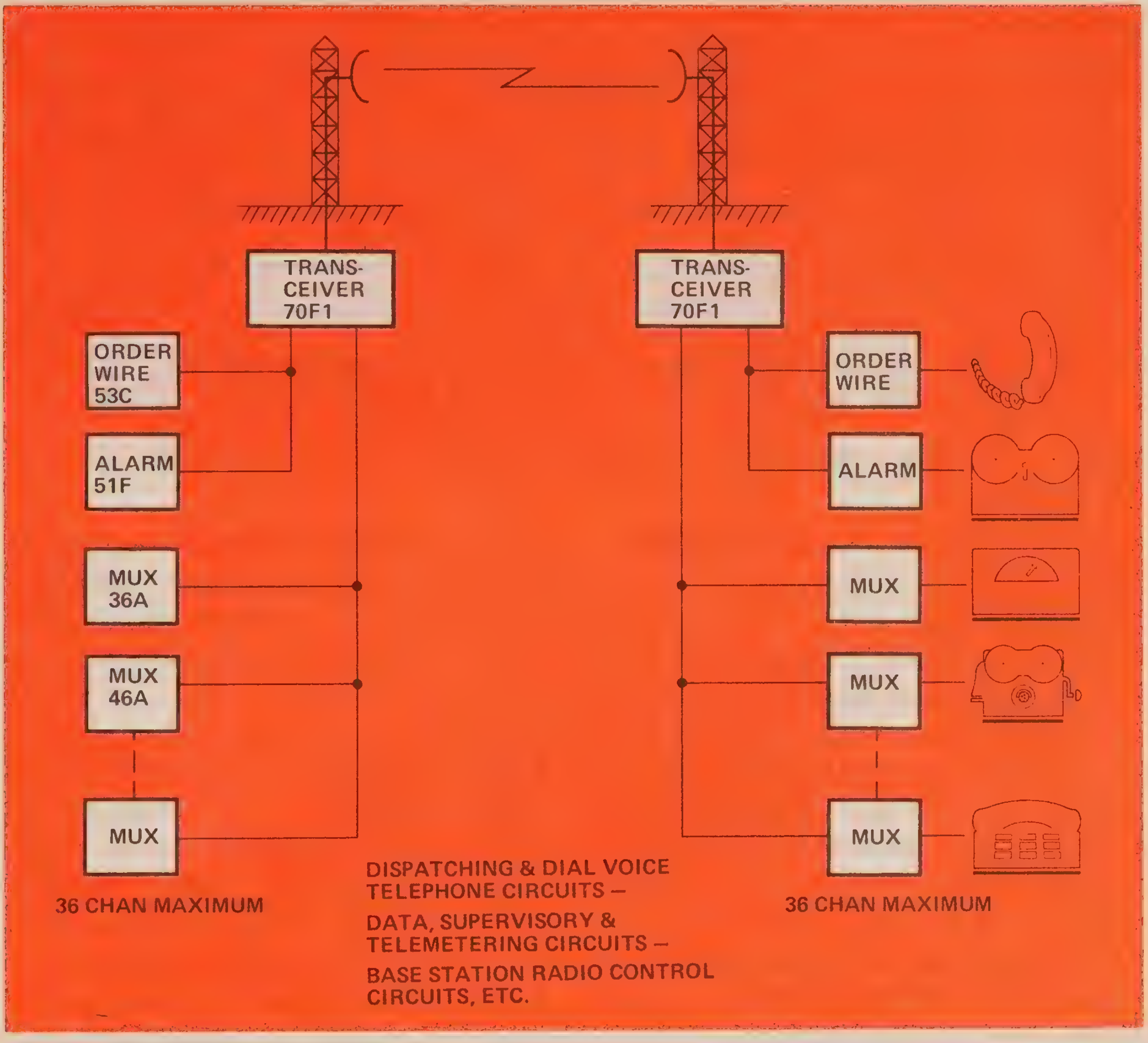


Figure 2. Typical layout of a light-route radio system.

of petroleum in a sample drops to a predetermined value, production is halted at that well head, permanently or until further drilling locates a new source of fuel.

Along the pipeline itself, status monitoring and remote supervisory control contribute significantly to the total operation, since through these functions volumetric output, petroleum flow rate, pressure, and temperature are monitored constantly. In the remotest of areas, valves are opened and closed, and pumps are automatically activated or stopped with the assistance of the microwave link.

Light-route systems may also be used in conjunction with mobile radios. Police dispatchers, for example, can relay traffic information (via operational fixed, light-route radio) from mobile units to a master station or another substation. Power companies use light-route circuits for dispatching and administrative purposes, along with such functions as remote alarming, supervisory control, relaying, and telemetering.

The operation of today's modern railroads requires that information such as telemetering, CTC (Centralized Traffic Control), and hot box indications, as well as administrative and teletype circuits, be brought in from low-density areas by the most reliable and economic means possible. Railroad companies extensively use microwave systems to this end. The status of

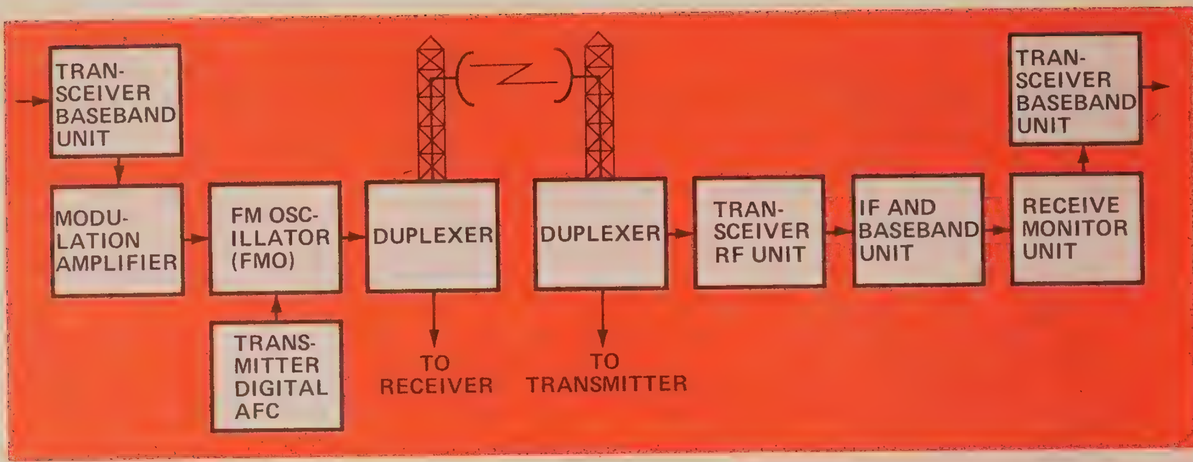


Figure 3. The basic units of the 70F1 microwave radio.

shipments and the locations of locomotives and freight cars is instantly available through the use of a microwave network tied to a central data processing center. Hot-box detection systems locate overheated journals (the rotating portion of a bearing) on moving trains. The hot-box detector transmits from remote sites, via microwave radio, information to a master station on the location of hot journals. This information is then transmitted by mobile telephone to maintenance crews along the line. An ideal use for a light-route system is in replacement of open-wire line along railroad tracks. Open-wire lines require a great deal of maintenance and expense that can be substantially reduced by use of a light-route system.

Another use of the light-route system is for telephone spur routes in a metropolitan area to reach small central offices. A tower in the center of the city provides a central transmission point from which several central offices can be serviced by light-route radios.

Three things of particular importance which a manufacturer must consider in the design of a light-route system are reliability, quality of performance, and economy. Adhering to these considerations, the GTE Lenkurt 70F1 radio, for example, contains such devices and techniques as a power fmo (frequency modulated oscillator), a digital afc (automatic frequency control), and a double-conversion receiver process.

FMO

The baseband information to be transmitted is applied to the transceiver baseband unit from multiplex equipment, or, in spur-route applications, from a backbone radio route (as when a few channels need to be branched at a repeater site). The transceiver baseband unit prepares the baseband signal, by way of level adjustments and filtering, for application to the modulation amplifier, where the conditioned signal is amplified in level to properly deviate the fmo (see Figure 3).

The fmo is a power (1 watt minimum) oscillator which provides a 2-GHz frequency modulated signal at its output. This means that the information contained in the original 36 voice channels is translated into a signal in the 2-GHz range which is of constant amplitude but of varying frequency. The fmo is frequency modulated by a varactor diode (a capacitance-varying device) which is coupled to the coaxial cavity resonator. The cavity determines the carrier frequency, and the variation in varactor

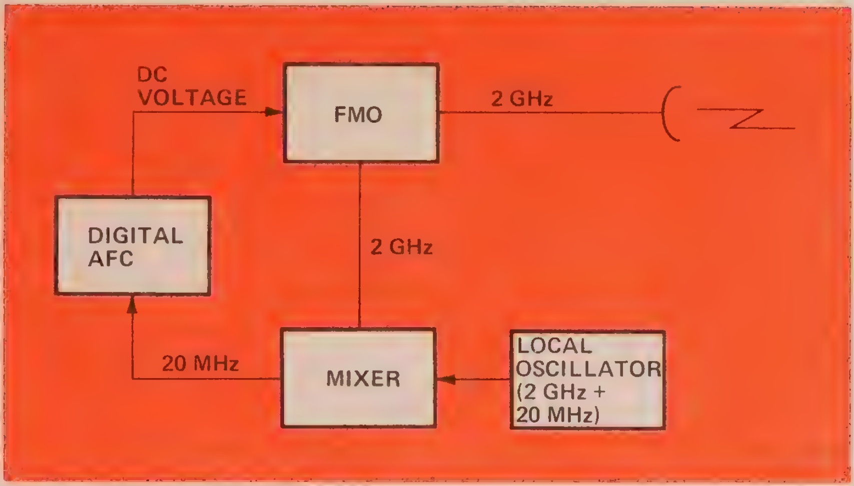


Figure 4. The digital afc controls the frequency of the fmo.

capacitance, due to the ac modulating signal, produces the frequency-modulated 2-GHz signal. This arrangement, in which the fmo produces a high power fm signal, eliminates the need for the conventional power amplifier, thus considerably reducing the cost of the system.

Digital AFC

The function of the digital afc is to control the frequency of the fmo. It does this by working within a closed transmitter loop as shown in Figure 4. The dc output from the afc drives the varactor in the fmo; the variation of dc voltage on the varactor varies the output frequency of the fmo so that it remains on the required 2-GHz frequency.

The 20 MHz from the mixer (see Figure 5) is amplified to a level sufficient to operate the digital IC's (integrated circuits). The signal is passed through a bandpass filter which allows only a band of frequencies centered around 20 MHz to pass. This prevents false locking of the afc on spurious tones such as 10 MHz or 40 MHz. The amplified 20 MHz sine wave is clipped so that distinct pulses are formed. These pulses are input to a counter, which is a string of IC's whose function is to count the number of input pulses. Each time 4096 pulses are counted, one pulse is generated by the counter. Effectively, this means that the number of input pulses is divided by 4096. The output from the counter, after processing the 20 MHz input

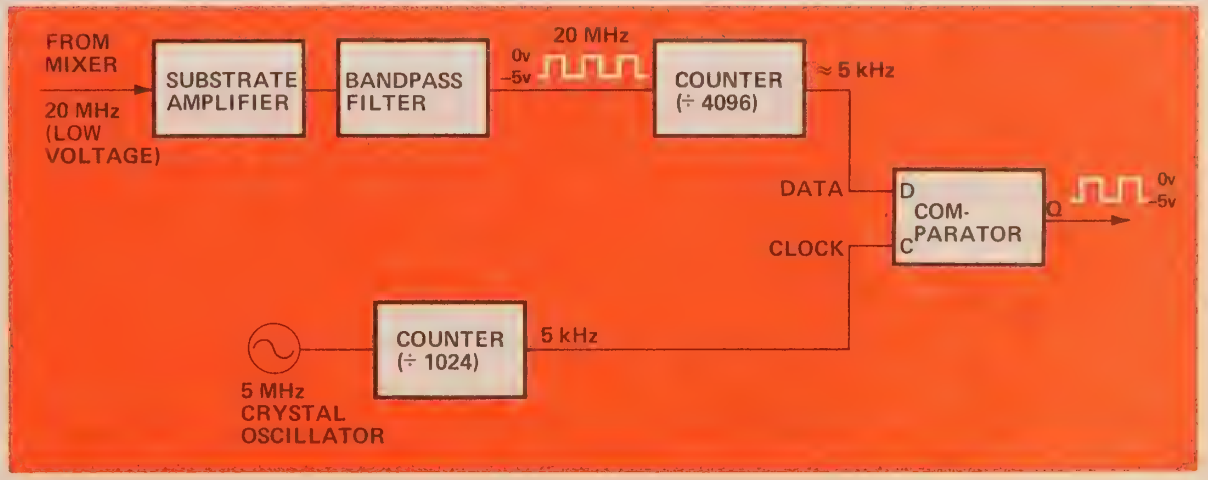


Figure 5. The basic components of the digital afc.

signal, is approximately 5 kHz, and is applied to one lead (data or D lead) of a comparator.

In another circuit, the output of a 5 MHz crystal oscillator is also fed into a counter which yields one pulse per each 1024 pulses of input. For a 5-MHz input from the crystal oscillator, an output pulse frequency of approximately 5 kHz is obtained. This output is fed to the second lead (clock or C lead) of the comparator. The output lead (or Q lead) of the comparator has as its output whatever binary state (high or low, one or zero, etc.) exists at the data lead when the clock pulse goes high. Each time the clock goes high, the output of the comparator will be whatever value is on the D lead. A zero on the data lead causes a zero output when the clock goes high. The output from the counter that is connected to the crystal oscillator is a constant frequency source. It is the 20-MHz output from the mixer where the variance in frequency will occur. If the mixer output were exactly 20 MHz, both outputs to the comparator would be the same. If, however, the output from the mixer were 21 MHz, the frequency from the counter would be slightly higher than 5 kHz (see Figure 6). The 21-MHz pulse stream will cause the input to the data lead of the comparator to rise sooner than would a 20-MHz pulse stream. The position in time (or phase) of the counted down output from the mixer at the time of sampling (which is determined by the output derived from the crystal oscillator source) determines whether the frequency is above or below 20 MHz. A frequency above 20 MHz will cause the comparator to give a high or 0v output, while a frequency below 20 MHz will give a low or −5v output. This output is amplified and integrated to yield a dc voltage. The variations in this voltage control the diode varactor in the fmo, which in turn keeps the output frequency stable at 2 GHz.

The Receiver

At the receiver, amplification is done directly at the input frequency by the rf pre-amplifier after being filtered by an rf bandpass filter. The pre-amplifier output is filtered by a second bandpass filter to suppress any amplifier noise generated on the image frequency (see Figure 7).

The filtered signal from the pre-amplifier is connected through a hybrid splitter to the balanced receiver mixer. The multiplied output of the LO/RO (local oscillator/reference oscillator) is injected into the mixer through the second port of the hybrid splitter. A three stage IF amplifier selects the 70-MHz product of the mixer and connects it to a second

Figure 6. The timing relationship of different signals as applied to the frequency comparator.

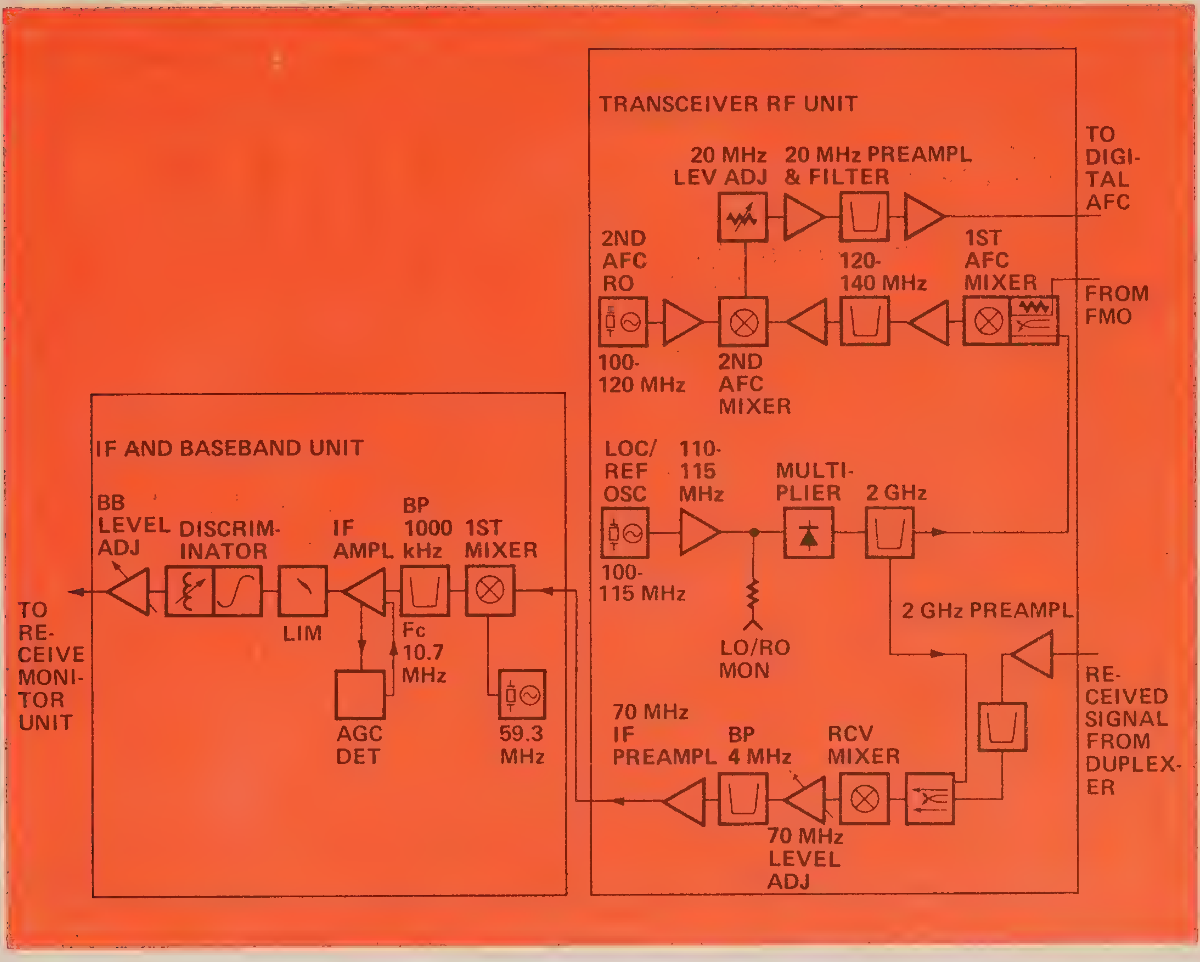


Figure 7. Amplification at the receiver is done directly at the input frequency.

receiver mixer (double-conversion process) in the receiver IF and baseband unit. The 70-MHz signal is mixed with the output of a 59.3 MHz crystal oscillator, and a bandpass filter selects the 10.7 MHz second IF signal. At 10.7 MHz, IF amplification is far easier and less expensive than it would be at 70 MHz, since transistor gain is quite adequate at this frequency, and the relative bandwidth of the filter for 36 channels is much wider (approximately 7 times wider) than it would be at 70 MHz. Also, the discriminator design is much simpler and easier to linearize at lower frequencies, and the complete circuit can be placed on thick-film substrates. The 10.7-MHz frequency modulated signal is demodulated in the discriminator and amplified to provide the necessary baseband levels.

Hot-Standby Protection

A protected system is one which has been provided with some type of security against total system failure. In many microwave systems this protection is in the form of "hot-standby" equipment which will take over the function of the main system should it fail. Protection for the GTE Lenkurt 70F1 transmitter is achieved by switching to a second radio through an rf coaxial relay, a technique that considerably simplifies the transmitter protection arrangement.

The transmit logic circuitry monitors the transmit pilot alarms, transmit afc alarms, and transmit power alarms from both transmitters. If an alarm occurs on the primary transmitter, the logic will cause the output to switch to the standby transmitter. If both transmitters are in an alarm condition, the

transmitter with the lowest priority alarm will continue to operate.

The receive logic circuitry monitors the receive pilot alarms and receive noise alarms. If a pilot alarm occurs, the logic will cause the receiver in alarm to be muted and the other receiver to be switched on line. If pilot alarms occur in both paths, no switching takes place.

A receive noise alarm occurs when the noise exceeds the preset level (nominally 58 dBrnc0). A receive noise alarm will cause the receiver in alarm to be muted and the other receiver to be switched on line. If both receivers have noise alarms, both receive paths will be muted.

Testing

Calibration of a microwave system is greatly simplified by use of a loop-back type of test set, such as the GTE Lenkurt Type 48615 Transceiver Test Set. This set enables one person to calibrate the system from one location rather than requiring one person at each location.

The first function of the test set allows a receiver deviation check. The receiver baseband level is set up with a calibrated 70-MHz frequency-modulated signal, which is used as a reference; transmitter deviation is then adjusted with reference to the calibrated receiver. In the loop-back function, the near-end transmitter transmits a signal into the test set where the rf frequency is shifted (usually by 50 MHz), and loops the signal back to the near-end receiver, where the proper deviation adjustments can be made for the system. The use of the test set is a cost savings, since one person can perform noise loading, set levels, and accomplish routine radio checks from one location.

The technology that has been developed and applied to the design of high-density microwave systems can also be applied to light-route radio systems. These systems can now provide the industrial and common carrier user with an economical form of communication which is both reliable and of high quality.

BIBLIOGRAPHY

1. Ball, G.V. "*The 48607 Transmitter AFC Unit (PSS 48607),*" San Carlos, California: GTE Lenkurt Inc., (November 1973).

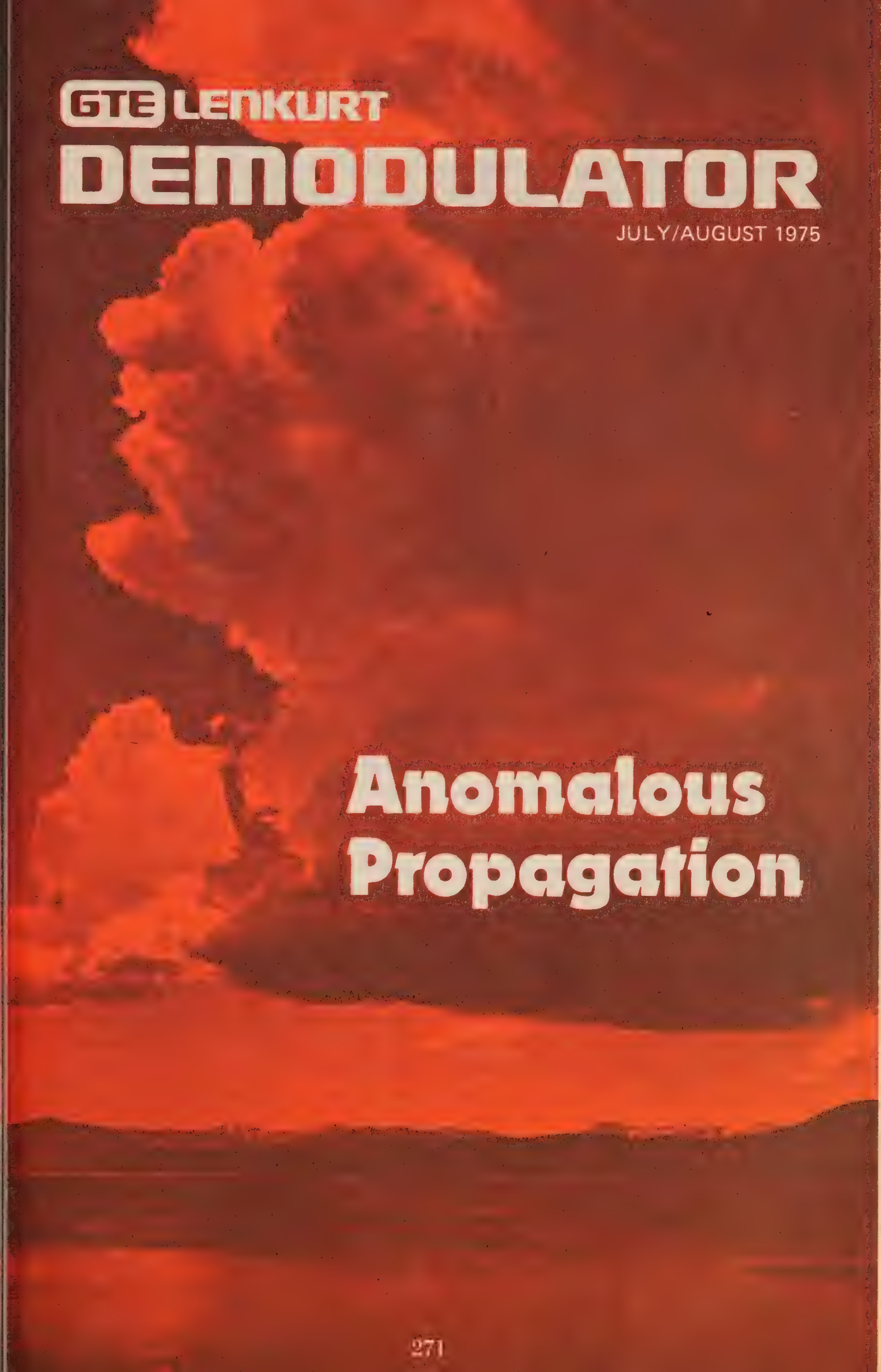
2. "Microwave Applications," *The GTE Lenkurt Demodulator*, May, 1972.

3. *70F1 Microwave Radio System, General System Practices (Section 394-706-110).* San Carlos, California: GTE Lenkurt, 1974.

GTE LENKURT

DEMODULATOR

JULY/AUGUST 1975



Anomalous Propagation

Anomalous Propagation – Part 1

Most microwave paths are subject to at least a small amount of multipath and attenuation fading throughout the year. This type of fading is normal and is taken into consideration in system design. Paths located in areas which support the formation of ground-based atmospheric layers may be influenced by a wide range of anomalous (abnormal) propagation effects, some of which defy traditional diversity or other conventional corrective techniques.

In its passage between transmitter and receiver, a microwave signal is subject to any number of obstacles that impede its progress to the intended destination. Many of these obstacles (such as buildings and mountains) are allowed for in the original design of the microwave path. Others, such as conventional types of atmospheric fading, are alleviated by system design and the use of diversity. There is, however, an area where conventional techniques normally used to combat fading are virtually ineffective; this is the area of abnormal, or anomalous, propagation, characterized in its most extreme form by the catastrophic phenomenon known as blackout fading.

Before launching into an in-depth discussion of anomalous propagation, it would be beneficial not only to discuss some of the terminology associated with fading, but also to review some of the characteristics, causes, and remedies of conventional fading phenomena. Since radio waves at microwave frequencies exhibit characteristics and properties similar to those of light waves, certain optical principles are useful in describing radio wave propagation; the most important of these are the principles of reflection, diffraction, and refraction. Individually, or in combination, these properties can significantly affect the reception of the microwave signal at the receiver, thus influencing overall per-hop or system reliability.

Reflection

Reflection of the microwave wavefront is a common source of received signal variation. Reflections occur when radio waves strike a smooth surface such as water, smooth earth, or the boundary between adjacent atmospheric layers of different densities. If both reflected and direct waves reach the receiving antenna, the ensuing cancellation may significantly reduce the received signal strength. Depending on the length of the reflected path compared to the direct path, the reflected wave may arrive at the receiving antenna either in phase, out-of-phase, or partially out of phase with the direct wave. Under conditions where the reflecting surface is very smooth and the reflected wave and direct wave are of equal amplitude and exactly out of phase at the receiver, the reflected wave may temporarily almost completely cancel the direct wave and cause a deep drop in received signal strength.

Diffraction

Microwave paths (and antenna heights) are ordinarily selected to pro-

vide line-of-sight clearance between the transmitting and receiving antennas. However, a direct path does not always guarantee good radio transmission. If the wavefront passes too near an obstacle, such as a hilltop or large building, the partial obstruction will result in an increase in transmission loss. If the microwave ray (the centerline route taken by the wavefront between antennas) is blocked (by inverse bending, for example), energy still arrives at the receiver by diffraction. The actual amount of obstruction or diffraction loss is dependent upon the Fresnel zone radius of the wavefront compared to the area obstructed and the reflectivity of the obstruction. The Fresnel zone radius at lower frequencies (VHF to 2 GHz, for example) is much larger than at higher microwave frequencies, and therefore sustains much lower obstruction or diffraction losses during inverse bending periods.

Refraction

Refraction occurs because radio waves travel at different speeds through media of varying density. In free space (a vacuum), the speed of a microwave signal is theoretically constant and maximum, but in the atmosphere, where the atmospheric density is higher due to the presence of gas and water molecules, the radio wave travels slower.

In a "standard" atmosphere, the pressure, temperature, and water vapor content (humidity) all decrease linearly with increasing altitude. Atmospheric density, the single parameter which combines the resultant effect of these variables, also decreases with altitude. Radio waves passing from dense air to thinner air undergo a change of direction in proportion to the difference in densities, since the portion of a wave front that enters the thinner air layer first begins to travel faster than the portion still in the dense air. The result of this process is a microwave path that is bent or refracted towards the denser atmosphere. In a uniform atmosphere, where the change in density is gradual, this bending or refraction of the radio wave is essentially continuous, so that the microwave beam is gently curved away from the upper to the lower atmosphere. The beam then generally tends to follow the curvature of the earth. This causes the radius of the earth to appear to the microwave beam to be larger than the true radius. When an extreme drop in atmospheric density with height (a negative refractive index gradient) occurs, or when the gradient is positive, climatic conditions are conducive to anomalous propagation.

The k Factor

The amount and direction of bending undergone by the microwave beam is defined either by the refractive index gradient or, more often, by the effective earth's radius factor, k. This factor, multiplied by the actual earth radius, gives the radius of a fictitious earth curve. The curve is equivalent to the relative curvature of the earth; that is, it is equal to the actual earth's curvature minus the curvature of the beam of microwave energy. Any change in the amount of refraction caused by atmospheric conditions can then be expressed as a change in k.

During standard atmospheric conditions, the range of k is from 1.2 in dry, elevated areas and 4/3 in typical inland areas, to 2 or 3 in humid coastal areas. The earth appears to become increasingly flat as the value of k increases (see Figure 1). When k equals infinity the earth appears to a microwave beam to be perfectly flat since the beam curves at exactly the same rate as the earth. If the value of k becomes less

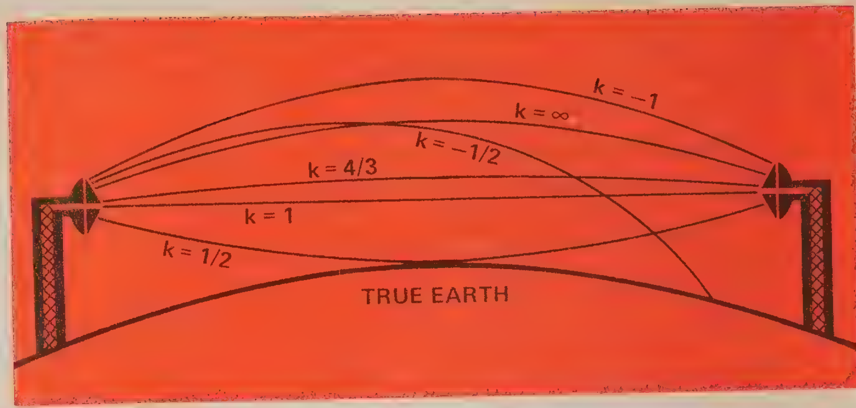


Figure 1. The degree and direction which a microwave beam bends can be defined in terms of the equivalent earth radius factor, k.

than one, the beam curves upwards or opposite to the curvature of the earth; to the microwave beam, the earth appears to "bulge." The effect of earth's bulge may be to partially obstruct the transmission path, causing an obstruction or diffraction. The curvature for various values of k can be calculated from the relationship:

$$h = \frac{d_1 d_2}{1.5k}$$

where h = the change in vertical distance from a horizontal reference line, in feet,

d_1 = the distance from a point to one end of the path, in miles,

d_2 = the distance from the same point as above to the other end of the path, in miles, and

k = the effective earth's radius factor.

Exempting multipath fading (easily overcome with diversity techniques), changes in k from 1 to infinity have little influence upon the received signal level of a properly engineered microwave path. Anomalous propagation occurs outside of this "normal" range of k. With k less than 1, the path could become obstructed and vulnerable to extreme multipath fading. When negative values of k occur, the path may become trapped and susceptible to blackout fading.

Superstandard Refraction

Superstandard refraction, (also called superrefraction), results from such meteorological conditions as a rise in temperature with increasing height (temperature inversion) or a marked decrease in total moisture content in the air with increasing height, either of which will cause a reduction in the atmospheric density with height. Under these circumstances, k increases, resulting in a flattening of the effective earth's curvature.

One of the conditions which may cause this type of abnormal refraction is the passage of cool air over a warm body of water. Evaporation of the water will cause an increase in humidity, and the low temperature near the surface is a sign of a temperature inversion. Low temperatures and high humidities greatly increase the atmospheric density near the surface, causing an abnormally high downward bending of the wavefront. In moderate instances of superrefraction, k approaches infinity and a microwave beam which is propagated parallel to the earth will remain parallel until obstructed or otherwise attenuated. An extreme case of superrefraction will bend the wavefront downward with a radius smaller than that of the earth (negative k), causing a blackout fade if the receiver is beyond the point at which the wavefront refracts into the ground (the radio horizon).

Substandard Refraction

Substandard or less than standard refraction occurs with certain meteo-

rological conditions which cause the atmospheric density to actually increase with height. This condition, described earlier as earth's bulge or inverse beam bending, causes an upward curvature of the microwave beam as shown in Figure 1, where k = 1/2. A substandard atmospheric condition may occur through the formation of fog created with the passage of warm air over a cool air or a moist surface. This will cause the atmospheric density to be lower near the ground than at higher elevations, causing an upward bending of the beam.

Comparison of Common Fade Types

Fading, the variation in the strength of a received microwave signal, is usually due to atmospheric changes and ground and water reflections in the propagation path. Figures 2 and 3 show selective and nonselective fading types that can influence the propagation reliability of a line-of-sight microwave radio system. Two or more types may occur simultaneously, complicating the received rf signal level as shown by the level recorder pattern. For example, rapid multipath fading often accompanies most other types of fading (except rainfall and blackout).

Fade types 2a through 2e result from one or more interference rays arriving at the receive antenna in and out of phase with the desired beam. These types of fades are selective, in that different frequencies within the transmitted band are affected differently; any type of diversity (frequency, space, hybrid, etc.) is almost totally effective in eliminating baseband hits or outages if adequate fade

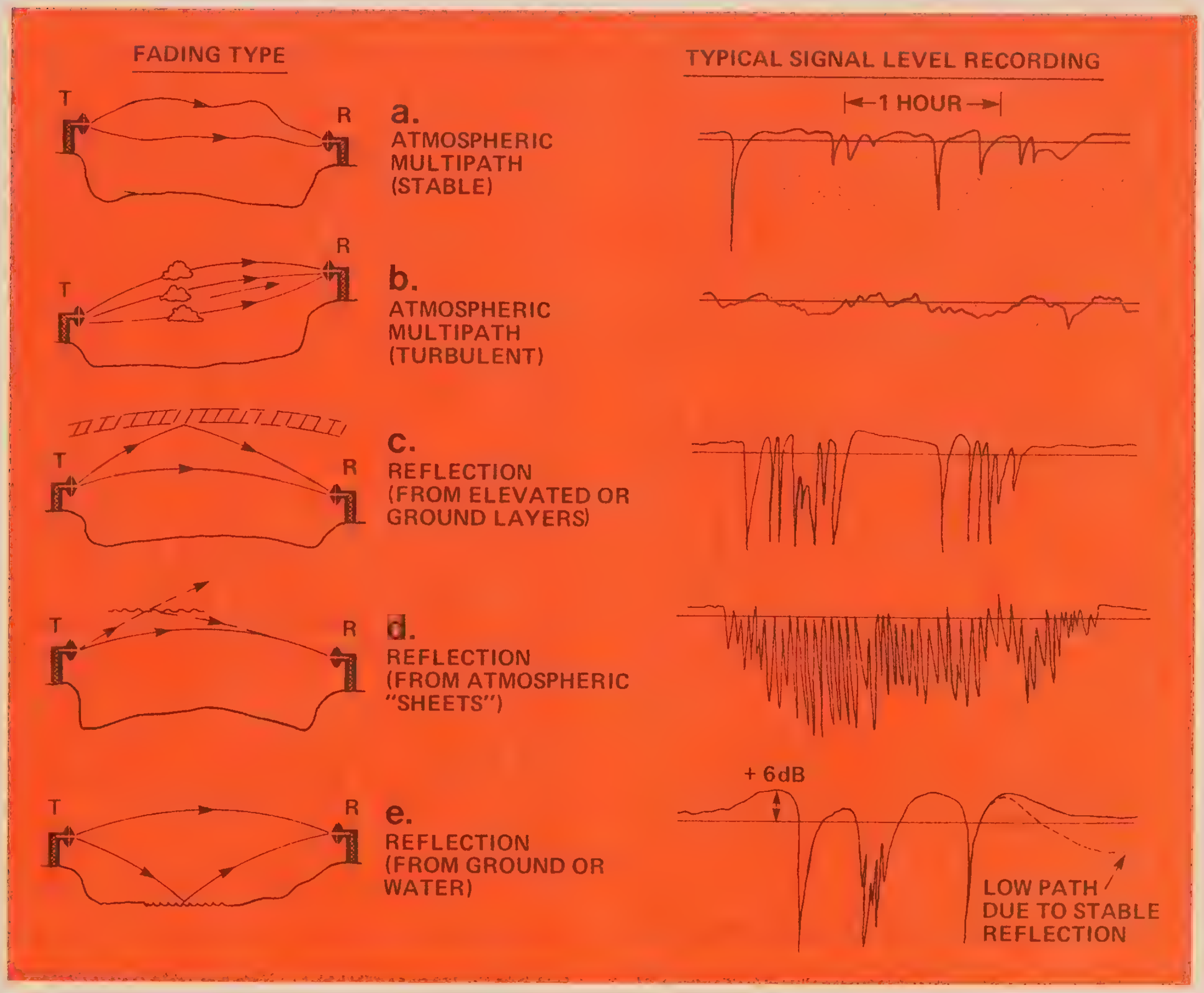


Figure 2. Selective fading types.

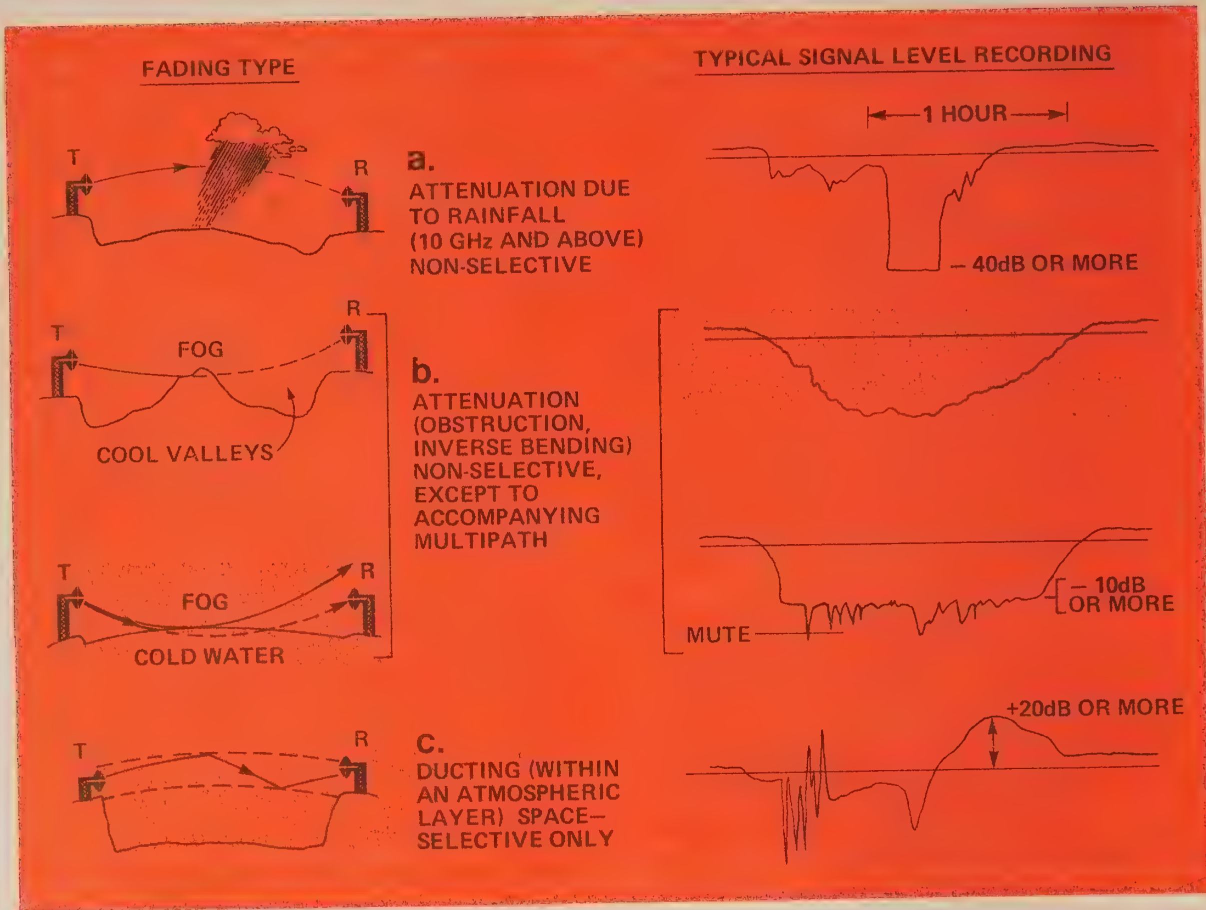


Figure 3. Non-selective fading types.

margin is provided. The power or attenuation fade types 3a through 3c are generally non-selective; normal frequency and space diversity techniques provide little or no improvement over a nondiversity configuration in propagation reliability.

Atmospheric multipath fading appears in both stable and turbulent form. Stable multipath fading (Figure 2a) occurs when only a small number (perhaps two or three) of secondary reflected or refracted paths (from the atmosphere or multiple low-amplitude ground reflection points) are received simultaneously with the desired path. The resulting fading characteristic is relatively slow, but occasional fast, deep fades can occur. Turbulent multipath (Figure 2b) causes fast but low-magnitude fades with fewer outages. Both types of atmospheric multipath fading have a time-versus-depth fade distribution; a 40-dB fade margin in poor propagation areas, for example, could approach about one hour per year of total outage time. Increased fade margin reduces outage time: for each 10-dB increase in fade margin, the outage time drops by a factor of about 10. Multipath fading is highly selective, and reduces baseband outage time by a factor of 100 or more. Such fading is also quite sensitive to antenna orientation and antenna size (larger antennas have better selectivity against slightly off-path secondary multipath rays).

Reflections from boundaries between elevated atmospheric layers and sheets are shown in the fade charts of Figures 2c and 2d. Elevated layer boundaries resulting from sharp changes in either temperature or humidity (or both) may form nearly perfect reflection planes at microwave

frequencies. Layers are usually quite stable, perhaps moving slowly, in a vertical direction; in- and out-of-phase reflections occur as this layer moves. Sheets (high-altitude, undulating layers perhaps 10 feet thick and 6 miles long) are constantly changing, and the wild, fast, deep fades (Figure 2d) occurring at night are often attributed to them. During severe reflection fading, the direct path is usually depressed by obstruction or antenna decoupling, increasing its susceptibility to the interference ray. Larger antennas and vertical polarization may help in such cases. Any type of diversity (the greater the spacing the better – 2% frequency or 40-foot antenna spacing are common) essentially eliminates baseband outages caused by reflections from layers and sheets.

Specular reflections from ground or water may occur as shown in Figure 2e. A stable signal reflected from a reflection point may combine with the desired signal either to increase the receive level as much as 6 dB, or to completely cancel the desired signal, perhaps causing a 50 dB or greater fade. In a varying atmosphere, the receive level chart displays a rolling pattern. A path "hung-up" on a stable reflection may remain depressed 10-20 dB for days or weeks. Increased fade margin is of some benefit, and the reflection point could be shielded by planting trees, by erecting a screen, or by relocating the antennas. Antennas can be tilted slightly upward to provide increased discrimination to the reflected ray. Larger antennas with smaller beamwidths are useful, especially at the low end of a high/low path (where one antenna is considerably elevated above the other). Diversity essentially eliminates outages caused by this type of fading, and optimum antenna spacings may be computed for a given path. Vertical polarization is also effective if the reflection (grazing) angle is larger than 0.1 degrees.

Attenuation due to rainfall is shown in Figure 3a. Although fades caused by rainfall are occasionally observed at lower frequencies (10-20 dB fades at 6 GHz have been recorded), this type of fade generally causes outages only on paths above 10 GHz. The outages are usually caused by blockage of the path by the passage of rain cells (thunderstorms, etc.), averaging 5 miles in diameter and 5-15 minutes in duration. The fading chart shows fairly slow, erratic level changes, with rapid path failure as the cell intercepts the path. The fades are nonselective (all paths in both directions are affected simultaneously). Vertical polarization appears less susceptible to rainfall attenuation than horizontal at certain frequencies. Increased fade margin is of some help in rainfall attenuation fading; margins as high as 45 to 60 dB have been used in some highly vulnerable 11-13 GHz links for increase reliability. When permitted, crossband diversity is totally effective – the lower 6-GHz path is stable (affected only by turbulent multipath fading) during periods when the upper 11-GHz path is obstructed by rain cells. Route diversity (paths separated by more than five miles) has also been used successfully. In some 13-GHz intercity video systems, a direct VHF link (receiving a weak but usable direct TV signal) is automatically switched on upon loss of the microwave link during rainstorms.

Attenuation fading due to partial path obstruction in a substandard atmosphere is nonselective. Increased fade margin and antenna heights are the prime solutions, although outages are usually caused by the accompanying severe multipath fade and not by the diffraction fade alone. The lower

frequency bands (2 GHz) exhibit less obstruction or diffraction loss than the higher bands for a given amount of inverse bending, so they are recommended where tower heights and fade margins cannot be increased.

Ducting

Paths trapped in ground or elevated ducts may exhibit wide, slow level changes, up to 30 dB above median in some instances, and thus may be compared to a gigantic waveguide coupling the transmit and receive antennas (see Figure 3c). Because ducts are often narrow, space diversity is sometimes helpful in reducing their effects. Frequency diversity is of little use except as protection against the accompanying multipath fading. In extreme cases of ducting, the microwave path may have to be rerouted around the ducting area, but outages caused by this mechanism are rare.

Blackout Fading

The occurrence of blackout fading is rare compared to other types of fading, but when it takes place, its effects are total and catastrophic. Traditional schemes used to improve the reliability of a microwave link, such as increased fade margin and conventional diversity, have only minimal influence with this type of fading.

Catastrophic path propagation failures due to blackout fading have been reported for more than 30 years. Some investigators call any lengthy depression of the microwave signal a fadeout, but this does not clearly identify the specific fade mechanism involved. For example, a non-outage fadeout could result from the partial obstruction of a path during substandard propagation periods. The terms ducting, trapping, blanking, drop-out, antenna decoupling, defocussing, and space-wave fadeout have also been adopted in describing various blackout events, but these terms could also be associated with phenomena other than the specific ground-based, superrefractive gradients (sharply decreasing density with height) causing the true blackout fade.

Common Denominators

Although propagation blackout fades occur in widely separated parts of the world, most have the following unique properties in common:

(1) The propagation failure is absolute, and no reasonable increase in fade margin can adequately resolve it.

(2) The microwave links are close to or traverse areas characterized by shallow, standing bodies of warm water such as swamps, shallow bays, and irrigated farmland, and may be located parallel to coastlines.

(3) The weather just prior to the blackout fade is unseasonably warm and humid; the fade coincides with a marked change in temperature and increase in humidity. The fades usually take place in the late evening, although some occur during daylight hours with the passage of a cold front.

(4) Most of the path lengths are in the 20-30 mile range, and, although antenna heights adequate to prevent a complete diffraction outage during subrefractive periods are usually provided, most have less than 150 feet of clearance over water or moist ground during normal propagation periods.

(5) The catastrophic outages occur simultaneously in both directions of transmission and in both diversity paths (except that in rare instances space diversity receivers showed that the fade exhibits some height selectivity).

The received signal level recording during a blackout fade shows a unique

pattern, different from any other fade type (Figures 4a and 5). All other types (selective or nonselective) will show widely varying fade depths from path to path and from day to day, depending upon terrain reflectivity or water roughness, rainfall intensity, path clearance, amount of obstruction, atmospheric turbulence, or the transmitted frequency. Blackout depends only on the thickness, refractive index gradient, and the frequency of occurrence of the ground-base atmospheric layer which, in turn, establishes whether or not the receive antenna is within the radio horizon of the transmit antenna. Blackout is generally a go/no-go phenomenon: either the path is well established or it is totally lost. Except for occasional changes in the thickness or refractive index gradient during blackout that may cause wide swings in received level, the fade depth is nearly always far greater than any reasonably assigned fade margin.

Blackout is most often confused with the diffraction (obstruction) fade that takes place with inverse bending (see Figure 3b and 4b), although the atmospheric refraction characteristics for these two fade types are reversed. Blackout fading results from the presence of an intervening superrefractive atmosphere, which is sometimes invisible except for boundary haze, and sometimes visible in the form of warm

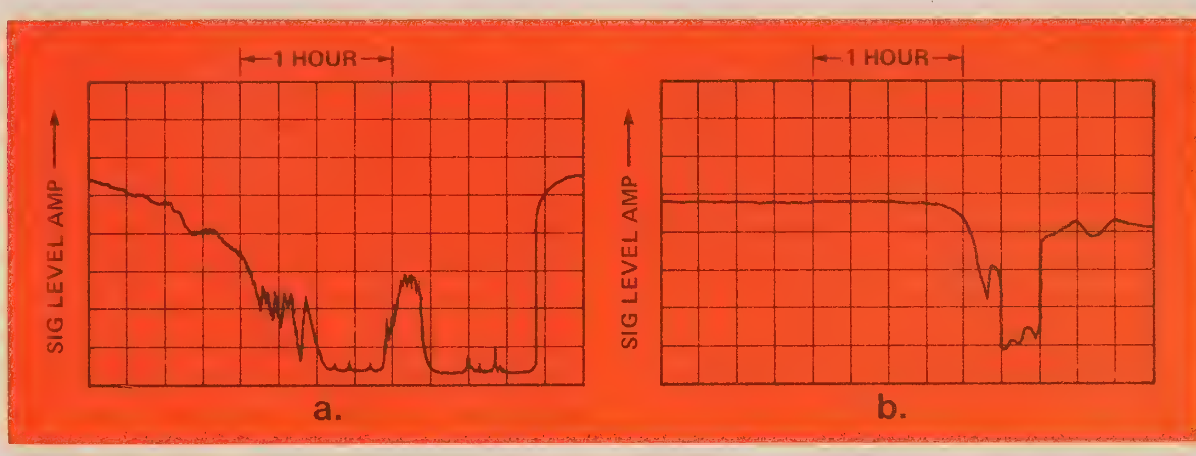


Figure 4. (a) Blackout fading on a low clearance path over shallow, warm water (superrefractive atmospheric layer). (b) Diffraction fading on a low clearance path over deep, cool water (subrefractive atmospheric layer).

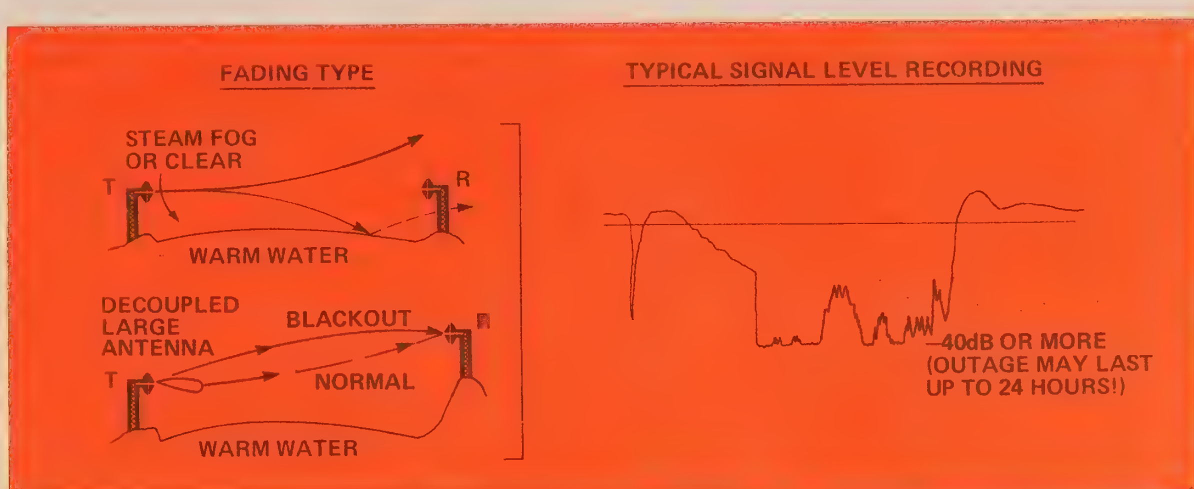


Figure 5. During blackout fading, the microwave path is trapped or arrives outside the main lobe of the antenna. This type of fading is usually non-selective, but may occasionally be space-selective.

water "steam" fog or mist that refracts the microwave wavefront downward or into the ground (or water) before it reaches the receive antenna. Usually, no part of the transmitted signal arrives at the receive antenna (see Figure 5). In contrast, inverse bending is the product of a subrefractive atmosphere (often cold-water fog) that may refract the wavefront upward before it reaches the receive antenna. But in this mechanism, a microwave path, although obstructed, continues to exist to the receive antenna, and a stable but depressed signal is still available, in contrast to the total outage typical of the blackout fade. Rapid multipath fading often accompanies the earth's bulge fade, so diversity protection is required in preventing baseband hits during this period. Again, in contrast, while the blackout fade often shows path instability, it is rarely accompanied by multipath fading, although some characteristics during blackout resemble multipath fading.

For example, if the radio horizon during blackout occurs just in front of the receive antenna, small changes in the thickness or refractive index gradient of the layer may cause the microwave signal to whip in and out of range of the receive antenna. The resultant fade pattern would be similar in appearance to interference fading with slightly varying path clearance with changes in k. But the blackout recording does not exhibit the 6 dB rolling upfade (increase in signal reception) characteristic of two-path reflection fading.

Multipath fading can precede blackout, but usually only over reflective terrain or water and under two exacting conditions. In the first instance, the direct path is partially obstructed by a low layer and a secondary superrefracted path reflected from the terrain arrives to cause interference fading. In the second case, the layer is unusually thick, and the two superrefracted rays arrive at the receive antenna as a direct path and a secondary reflection from the ground.

The Atmosphere

Blackout fading results from the formation of unusually steep, negative atmospheric density gradients; a dramatic drop in humidity, or an increase in temperature with height, for example. As the microwave beam is propagated through a blackout atmosphere, the lower part of the wavefront is slowed by the dense air with respect to the upper part, causing a tilting of the wavefront toward the ground or water. The amount of tilt is very small, but over a 20-mile distance the amount of accumulated bending could refract the propagated beam into the ground before it reaches the receive antenna. The beam may then be absorbed by foliage or crops, or scattered by rough terrain, but is most often specularly reflected by smoooth terrain or water (in which case it will reappear again at a greater distance).

Atmospheric density and refractive indexes are important in determining k factors for establishing path clearances. The density of the atmosphere, in terms of its refractive properties, is a function of pressure, temperature, and humidity in the following approximate relationship:

$$N = \frac{79P}{T} + \frac{3.8 \times 10^5 eH}{T^2}$$

where: N = Atmospheric refractive index (N-units)
P = Pressure, millibars
T = Temperature, °K (273 + °C)
e = Saturation vapor pressure, millibars
H = Relative humidity, %/100

The refractive index gradient dN/dh may be determined from the refractive index measured at the surface (N_S) through the relationship:

$$dN/dh = -7.32\ \epsilon^{.005577 N_s}$$

and:

$$k = [1 + (dN/dh)/157]^{-1}$$

The k factors most commonly used in microwave path engineering are listed in Figure 6. The surface refractive index is related to a reduced-to-sea-level value of refractivity in approximately the following manner:

$$N_O = N_S\ \epsilon^{h/7}$$

where N_O and N_S are the refractive indexes at sea level and on the surface, respectively, $\epsilon = 2.71828$, and h is the altitude above mean sea level (AMSL) of the surface, in kilometers. For example, $N_O = 301$ would equate to a standard refractive index, N_S, at Denver (elevation 1.63 km) of 239 N-units.

All of these relationships are shown in Figure 7. The wide range of refractive index gradients present in a ground-base atmospheric layer may be seen in the vertical profiles. The standard –40 N-units/km median gradient is typical of inland microwave path propagation in the U.S. With higher ground humidities, a steeper –58 N-units/km (k=1.6) is usual around the Gulf of Mexico and other coastal areas, while in dryer elevated regions (such as in the Rocky Mountain states) dN/dh is typically –30 N-units/km (k=1.25).

The seasonal and diurnal median values and ranges of N_O, N_S, and dN/dh that take place throughout the U.S. and around the world are found in books published by the Department of Commerce *(OT Report 75-59, Refractivity Gradients in the Northern Hemisphere, ESSA Monograph 1, A World Atlas of Atmospheric Radio Refractivity,* and *NBS Monograph 22, Climatic Charts and Data of the Radio Refractive Index for the United States and the World).* Charts are also provided that show the frequency of occurrence of ground-base trapping layers.

In addition to the refractive index contour maps of the U.S. and the world, some of these publications show the distribution of atmospheric gradients with time at selected locali-

k, EFFECTIVE EARTH'S RADIUS FACTOR	dN/dh, N UNITS/km	ATMOSPHERIC CONDITION	MICROWAVE PROPAGATION
5/12	+220	HUMIDITY INVERSION (LARGE POSITIVE GRADIENT)	EXTREME EARTH'S BULGE (DIFFRACTION FADE)
1/2	+157	MODERATELY SUBREFRACTIVE	MODERATE EARTH'S BULGE
2/3	+80	SLIGHTLY SUBREFRACTIVE	SLIGHT EARTH'S BULGE
1	0	HOMOGENEOUS	NO REFRACTION
1.25	–30	DRY ATMOSPHERE	STANDARD (MOUNTAINOUS)
4/3	–40	STANDARD ATMOSPHERE	STANDARD
1.6	–58	HUMID ATMOSPHERE	STANDARD (COASTAL)
∞	–157	MODERATE NEGATIVE GRADIENT	FLAT EARTH
–1	–314	STEEP GRADIENT	POSSIBLE BLACKOUT
–0.5	–470	EXTREME GRADIENT	BLACKOUT

Figure 6. Commonly used k factors.

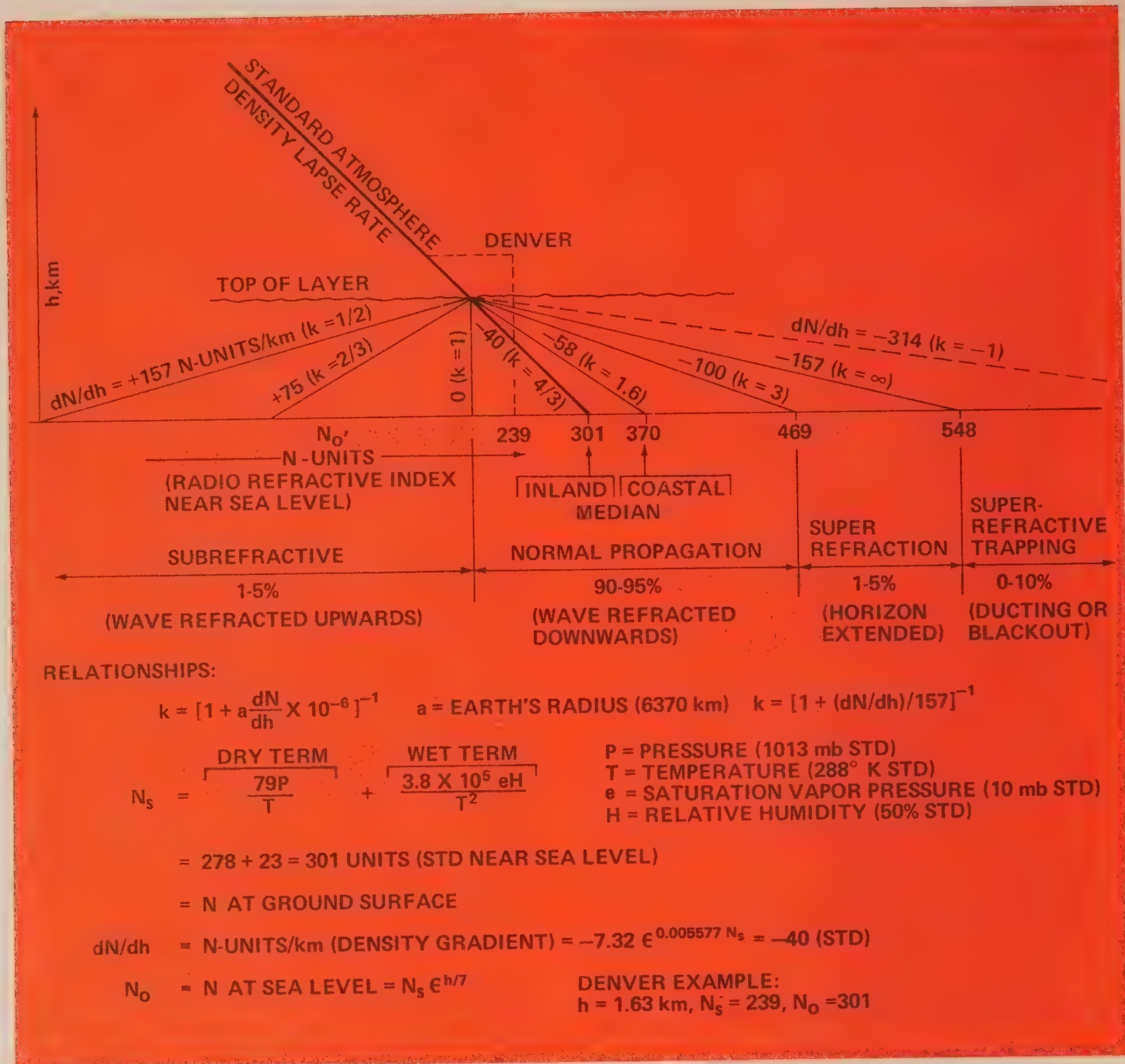


Figure 7. Atmospheric density profiles in subrefractive, standard, and superrefractive layers.

ties. Figure 8, a composite of many charts, shows the major components of a dN/dh (or k) probability distribution chart. In dry, inland areas the annual variation in k is small, with no abrupt changes in the distribution. In humid regions, a low-clearance path over moist ground or warm water is subject to a wide range of refractivity, from inverse bending (5% of the time, as shown on the left) to blackout (0.3 to 3% of the time, on the right). The occurrence of ground-base layering is immediately evident if a sharp bend in the probability curve occurs at the left (subrefraction or earth's bulge) or the right (superrefractive trapping, with possible blackout).

Positive dN/dh gradients (k less than 1) reduce (and possibly obstruct) path clearance, a frequent occurrence in paths traversing an expanse of cool, deep water. A dN/dh gradient of −157 N-units/km equates to k = ∞ (flat earth propagation). The range of gradients between −100 and −157, due to superrefractive atmosphere, simply extends the over-the-horizon range of a microwave system (unless obstructed by terrain), and has little effect upon the received signal levels on the established path.

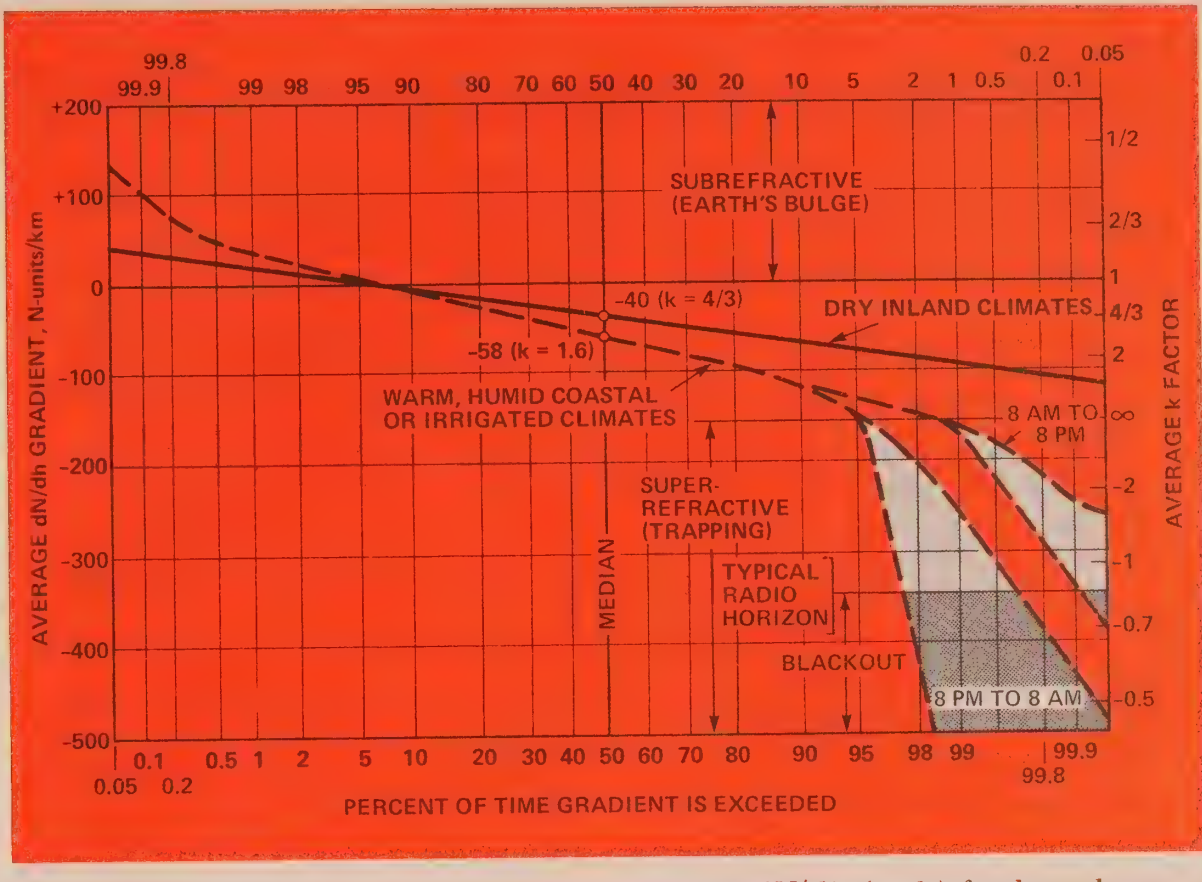


Figure 8. Cumulative probability distribution of dN/dh (or k) for low clearance paths (with layering) in dry and humid climates.

With radio beam curvature smaller than −157 N-units/km (negative k's), however, the microwave beam refracts downward faster than does the curvature of the earth and becomes trapped within the superrefractive duct. A blackout fade occurs if the microwave wavefront is refracted into the surface before it reaches the receive antenna.

Part 2 (the August, 1975, *Demodulator*) of this series will delve deeper into the characteristics and causes of blackout fading, and will suggest several remedies should this phenomenon occur.

BIBLIOGRAPHY

1. Laine, R.U. *Blackout Fading in Line-of-Sight Microwave Links.* San Carlos, California: GTE Lenkurt Inc., 1975.

2. Laine, R.U. *Determining Microwave Antenna Heights for Coastal Areas (Monograph 197).* San Carlos, California: GTE Lenkurt Inc., 1964.

3. Thomas, L.R. "Microwave as Applied to Railroad Operation in the Gulf Coast Area," *AIEE Communications and Electronics,* (January, 1974).

4. White, R.F. *Engineering Considerations for Microwave Communications Systems.* 2nd Ed. San Carlos, California: GTE Lenkurt Inc., 1972.

Anomalous Propagation—Part 2

Blackout fading, caused by signal trapping within a ground-based superrefractive layer, can introduce extended outages on low-clearance microwave paths. This rare but catastrophic type of anomalous propagation, which sometimes causes radar equipment to temporarily scan far beyond the normal radio horizon, is often the dominant factor limiting path length in microwave communications systems located in humid regions.

Part 1 of this series on anomalous propagation provided a general discussion of the characteristics of conventional fade types and their remedies, and on the nature of blackout fading. This issue will further investigate the phenomenon of blackout fading, discuss how it affects the microwave path, and what can be done to minimize its effects.

An atmosphere becomes superrefractive, and a low-clearance microwave path through it becomes susceptible to blackout fading, when cooler air passes over or develops above warm, moist soil or a warm body of water. This air mass may be produced by one of two processes: (1) with the passage of a cold front over the warm, moist ground or water at any time of the day or night, or (2) by subsidence.

Subsidence, derived from the word subside (to descend or tend downward), is the slow settling of a high-altitude, dry air mass from a high-pressure system. The air mass is warmed by adiabatic compression, overlaying and entrapping a cooler, moist air mass supported by surface moisture.

Superrefractive layers occur most often during the clear, calm, cool evening and early morning hours, but seldom appear during the afternoon, with its accompanying warmth, low humidity, atmospheric mixing and air turbulence.

Propagation Within the Superrefractive Layer

An analysis of the propagation of a microwave wavefront within a superrefractive layer is simplified if the wavefront is represented as a single line ray, the propagated path is presumed to be bilateral (the transmit and receive rays travel reciprocally on the same route) and it is assumed that the propagated ray will penetrate the layer, rather than being reflected from the boundary surface. Figure 1 shows a ray analysis of several antennas at various locations within a fully developed, ground-based superrefractive layer formed in a cool atmosphere over warm, shallow water.

In the instance where the transmit antenna is located above the layer (Figure 1a), two rays are shown, one above and one within the duct. The upper ray is propagated normally over a $k = 1$ to $k = 3$ range, depending upon the non-trapping refractive gradient above the duct. A receive antenna

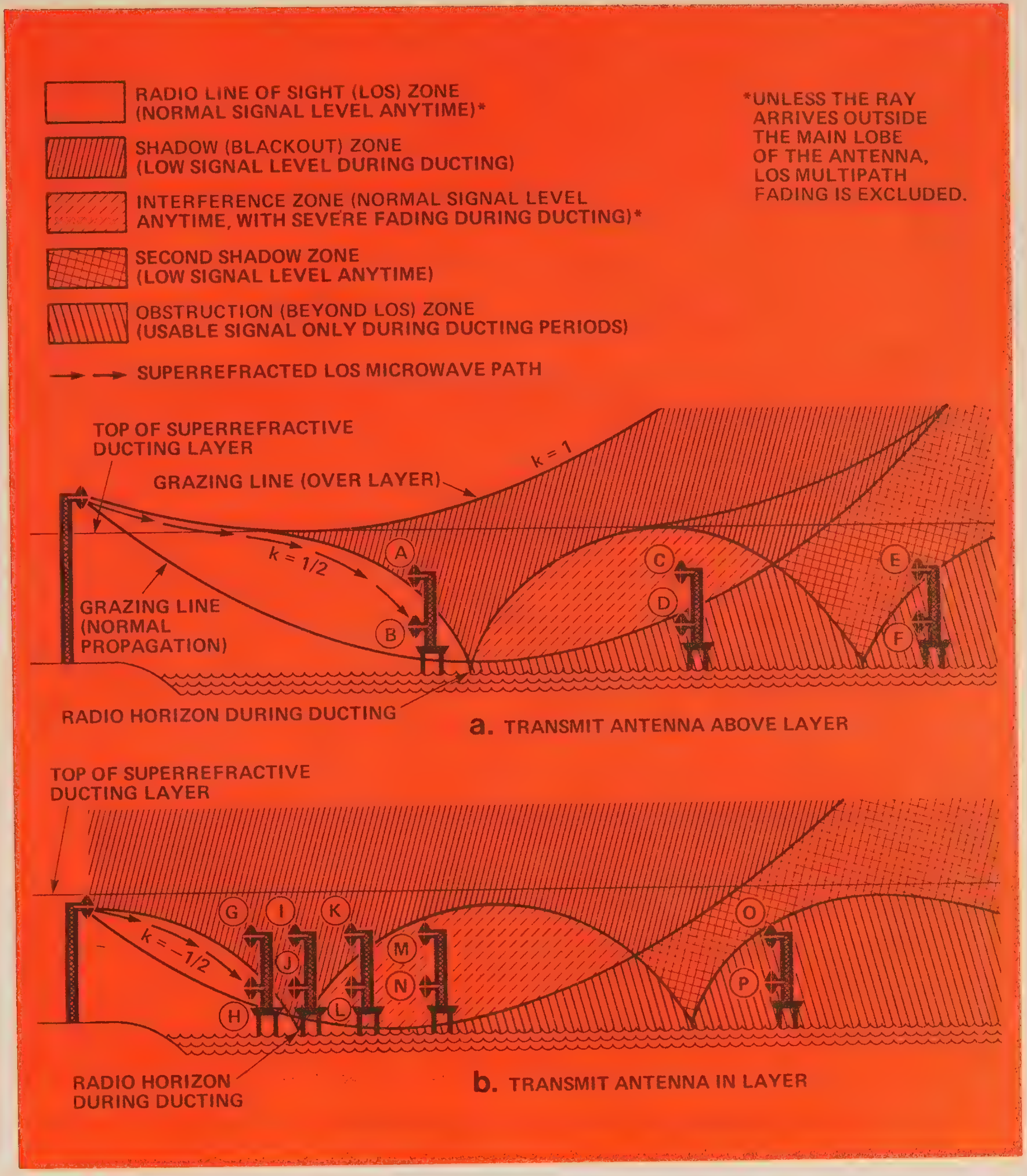


Figure 1. Tracing of a microwave beam within a superrefractive ground-based layer with (a) transmit antenna above layer and (b) transmit antenna in layer.

intercepting this ray would see a normal or perhaps partially obstructed signal level.

Within the duct, antenna B continues to see a high signal level if its radiation pattern is wide enough to avoid discriminating against the superrefracted incoming ray (dashed line), a factor in every superrefracted path. Antenna A, in the first shadow zone, is in blackout.

Antenna C was blacked out as the layer rose from the water, and recovered with a high but perhaps fluctuating signal level as the layer reached its present height. Antenna D, below line-of-sight and therefore obstructed during normal propagation, now has

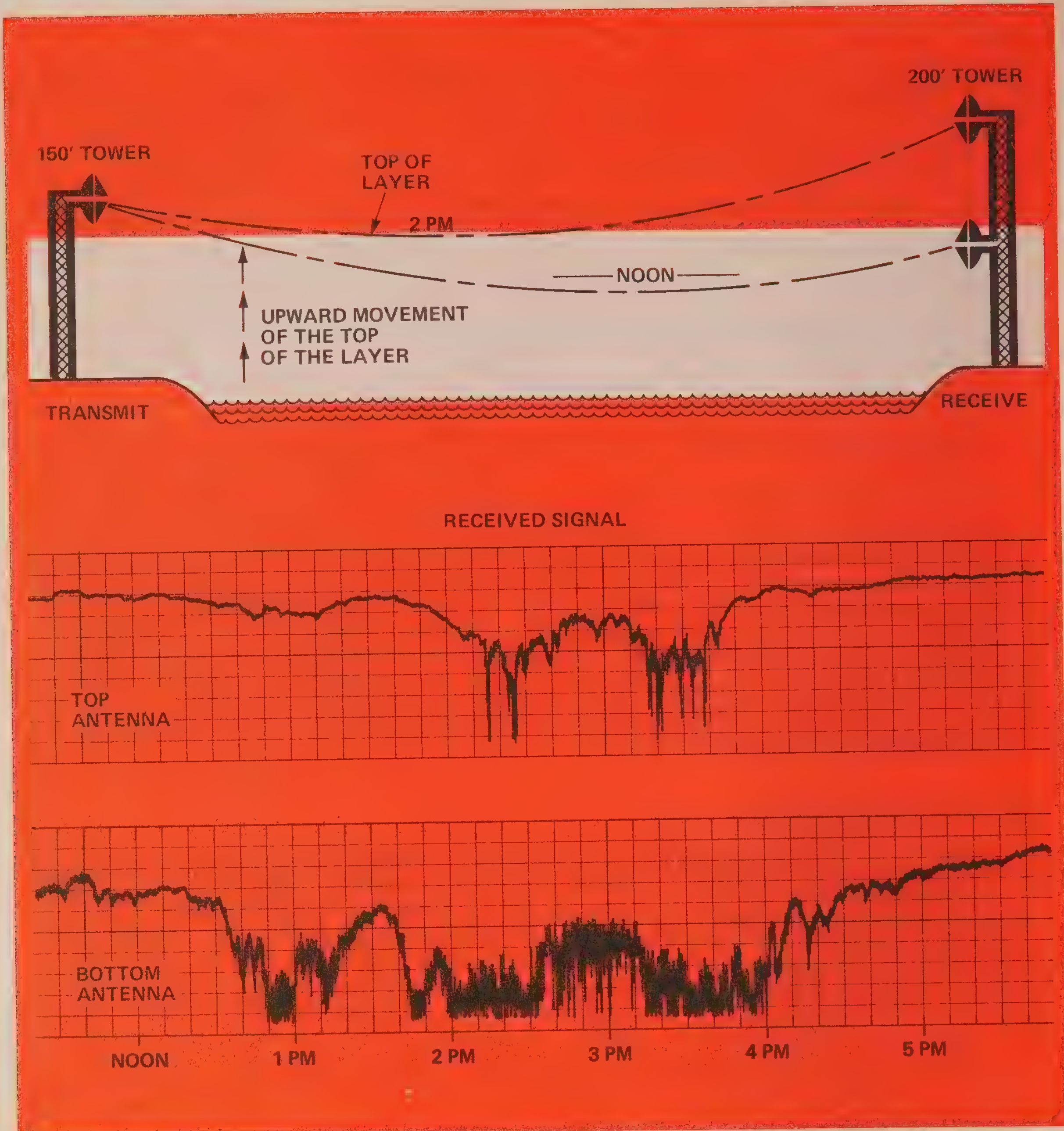


Figure 2. Blackout fading on lower space diversity antenna.

the same high but fluctuating signal level as antenna C.

Antenna E, far below line-of-sight during normal propagation and now in the second shadow zone, is still without input although its rf input fluctuated as the layer rose from the water. The received signal level at antenna F is normal as long as the layer exists.

The higher the transmit antenna, the more distant the radio horizon and range of transmission. If the transmit antenna is low, and within the layer (Figure 1b), the radio horizon range is shortened. Antennas G, I, J and K are blacked out; H has a normal input; L, M, and N have high but fluctuating inputs, and O and P (normally beyond the radio horizon) now also see normal but widely fluctuating signals. Direct superrefracted paths from the transmit antenna to antennas L, M, and N may exist if the layer is very thick.

Signal level recordings of diversity receivers with antennas spaced 100 feet apart are shown in Figures 2 and

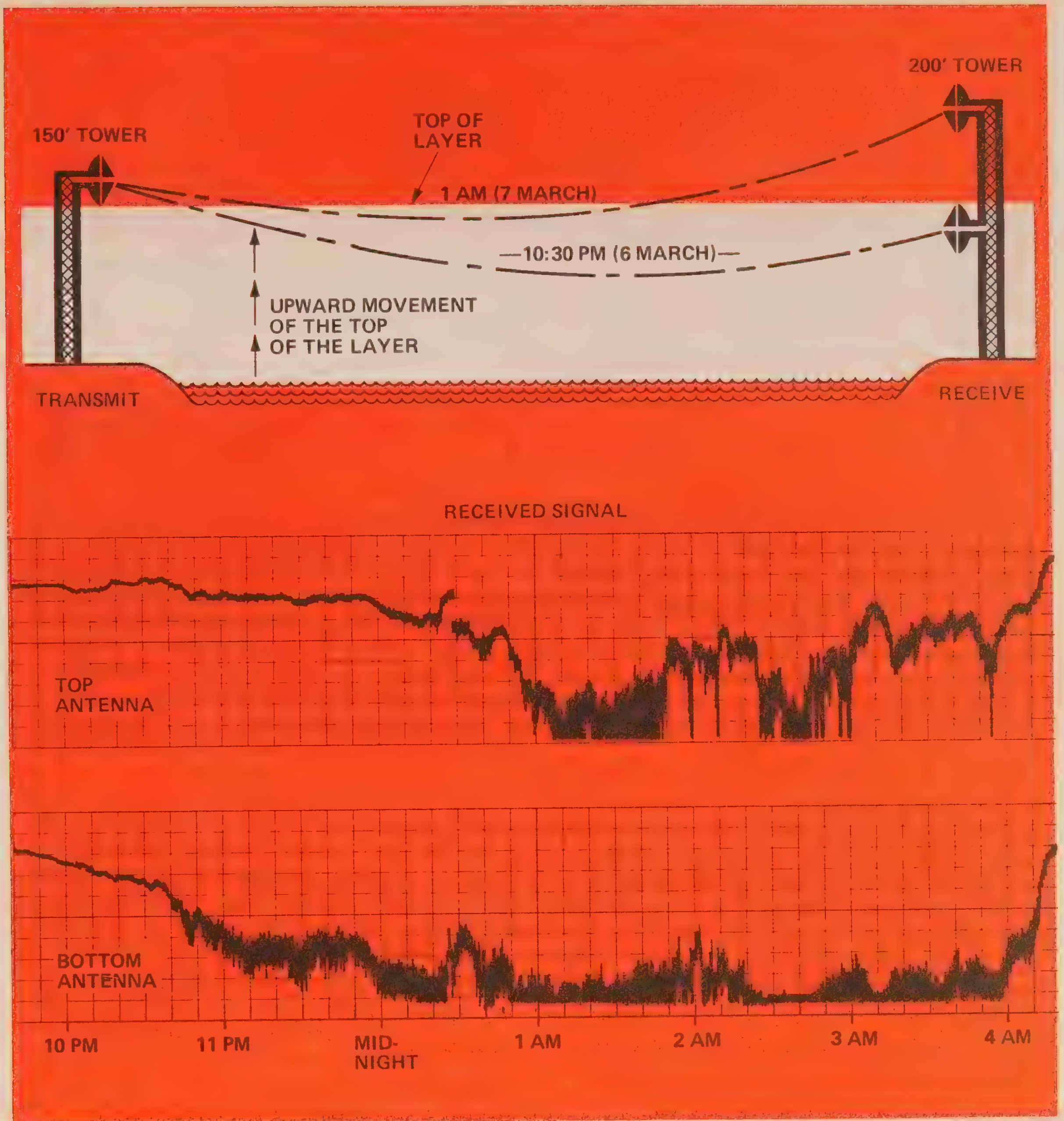


Figure 3. Blackout fading on both space diversity antennas.

3. In Figure 2, the layer moved upward to trap the lower path, which went into blackout (a rare daytime occurrence caused by the movement of a cold front into the area). The layer failed to rise above the path to the upper antenna, which experienced multipath fading, but not blackout fades (the path was reflective) commencing at shortly after 2 p.m.

On another occasion, the layer rose to a higher level and blacked out both paths, as shown in Figure 3. The upper path was out for 3 hours and the lower for more than 5 hours. The slow, strong, sporadic surges in signal level typical of the blackout mechanism are clearly visible in the recording, and are caused by slow changes in the refractive index gradient or thickness of a stabilized superrefractive layer.

Blackout Caused By Antenna Decoupling

Even if the receive antenna is within the radio horizon in the presence of

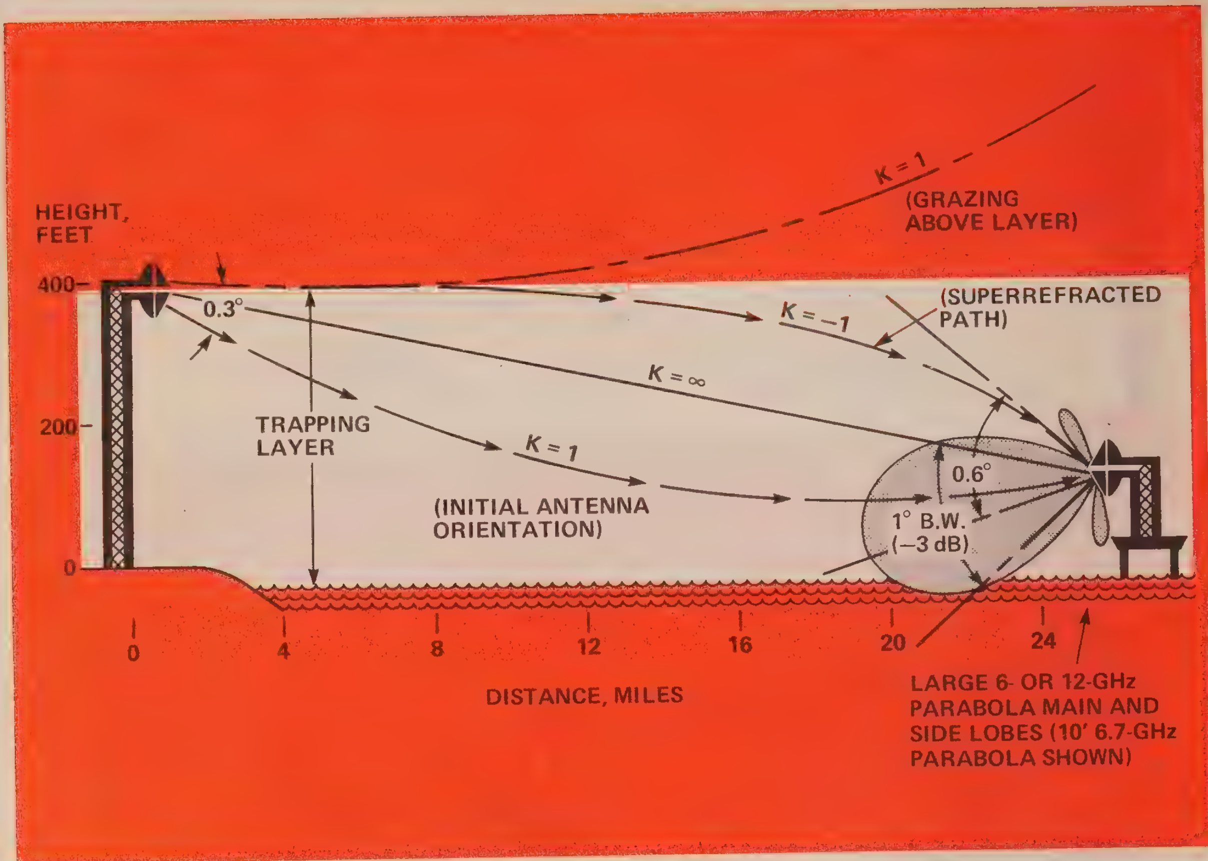


Figure 4. Antenna decoupling in a superrefractive ground-based layer.

a superrefractive layer, the propagated ray will arrive at a greater elevation angle than it normally would. If, during their initial orientation, the transmit and receive antennas were peaked for maximum signal during standard atmospheric conditions, a major upward change in the arrival angle could occur in a superrefractive atmosphere. With the narrow beamwidths common to large antennas or passive reflectors and long paths, a change in arrival angle of up to 0.5° or more may put the path outside the main lobe of the antenna, causing a blackout fade. Figure 4 shows the geometry of such an occurrence.

The characteristic behavior of such an antenna-decoupling blackout fade is identical to a blackout fade resulting from the receive antenna being beyond the radio horizon (see Figures 2 and 3). If antenna decoupling is anticipated (or experienced), the receive antennas may be tilted slightly upward (causing a 1- or 2-dB loss during normal propagation) or the lower antenna in a high-low configuration may be reduced in size. Antenna tilting has the further benefit of providing additional discrimination to a ground or water reflection during normal propagation periods.

Propagation Through a Partial Layer

If the microwave link is perpendicular to a coastline, or only a small part of the path traverses areas suspected of supporting a trapping layer, it is most unlikely that a blackout outage will ever occur. A superrefractive atmosphere capable of supporting steep, negative gradients seldom encompasses an area more than 2-5 km from the moist, warm surface generating such gradients (such as a shoreline). If the layer is close to only the high end of a

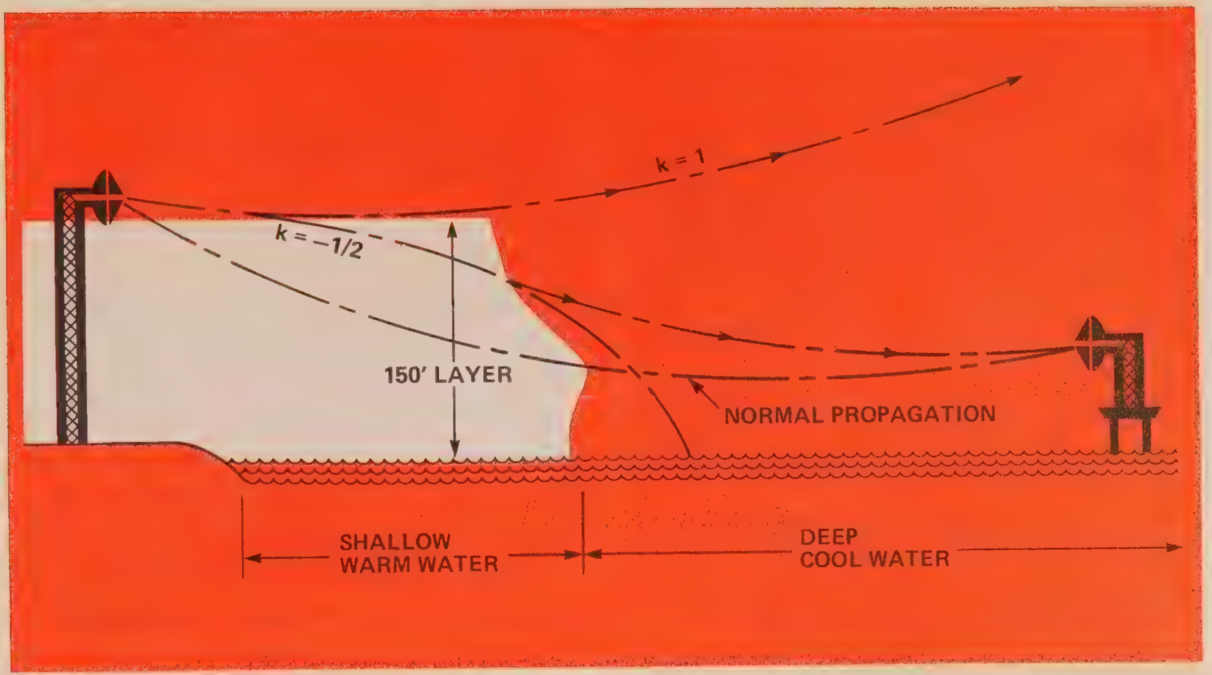


Figure 5. Ray tracing within a superrefractive layer enveloping only a part of the path.

high-low path, the possibility of trapping is greatly reduced as shown in Figure 5.

Low-clearance paths running parallel to a coastline or traversing a swamp or irrigated area at midpath or near the low end of a high/low microwave path would be most vulnerable to the blackout fading mechanism and should be avoided.

The Anatomy of a Blackout Fade

Propagation tests conducted with vertically spaced antennas have revealed that the superrefractive layer precipitating most blackout fades is ground-based (rather then elevated), and has its beginning at the moist ground or water surface and expands upward, rather than moving from a distant location.

A typical blackout fade sequence is illustrated in Figure 6. Six discrete stages or events are shown, any one of which may be more or less pronounced in any given system. The fade pattern often exhibits a high degree of instability during the blackout process.

The sequence is usually preceded with high ground-base water vapor content or high humidity (dew point very near the ambient temperature), and is triggered with a rapid change in either of these parameters (a 15° rise in wet bulb temperature, or a 10° drop in ambient temperature over a 3-hour period, for example). The superrefractive layer rising from the moist ground has a stable, clearly defined boundary from which a reflected ray could cause a deep interference fade in the system (event 2).

As the layer continues to expand in an upward direction, it becomes a path obstruction, causing a slow diffraction fade (event 3). When the layer actually intercepts the ray, the path is refracted downward away from the receive antenna, effecting a fully developed blackout fade as shown by event 4. The blackout may have a duration of many hours; slight changes in the refractive index gradient or thickness of the layer during blackout may result in slow, rolling increases in signal level (event 5). The layer then

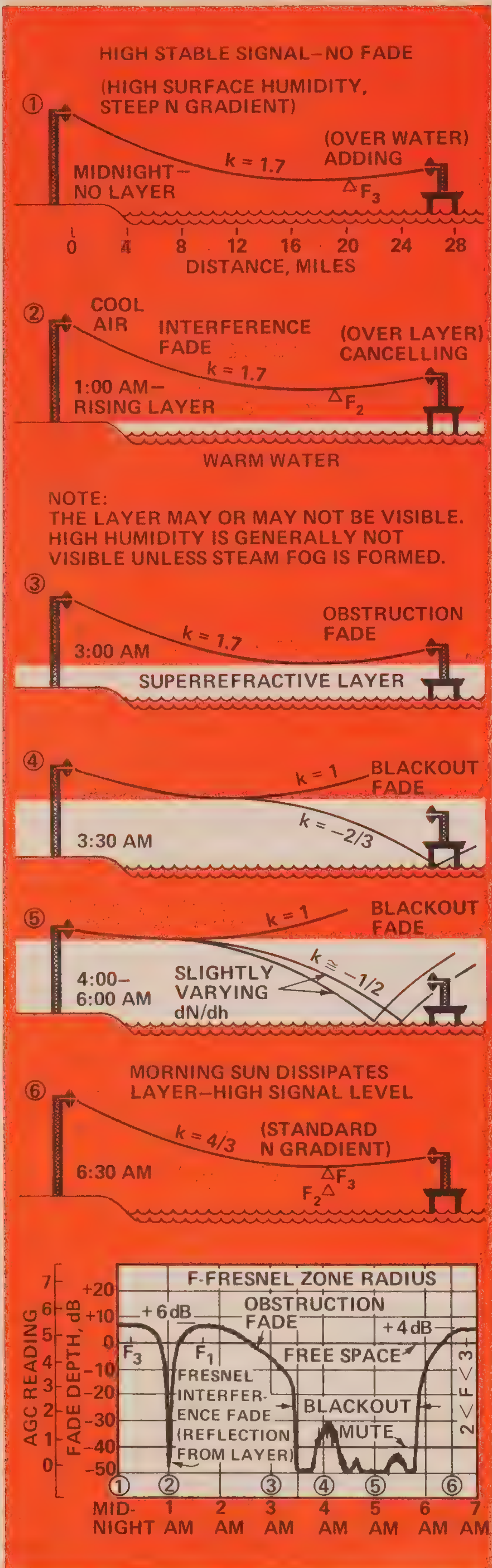


Figure 6. Blackout fade sequence of events.

dissipates with the warmth of the morning sun (event 6).

Outage Time With and Without Blackout

The fading distribution of a typical 22-mile, 6.7-GHz microwave path traversing areas having good, average, and poor propagation characteristics is shown in Figure 7. Any region supporting superrefractive layering of such severity as to cause occasional blackout is probably a poor propagation area whose multipath outages approach a Rayleigh distribution. The cumulative probability distributions with and without blackout are shown. Without trapping, the estimated outage time for the nondiversity system with 40-dB fade margin is about 19 minutes per year.

Blackout fading outage time was found to approach 44 hours per year on this path, the sum of about 11 events. The 11 blackouts, or periods of superrefractive trapping when the link (shown in Figures 2 and 3) exhibited sufficient fade instability as to be unusable, lasted from 20 minutes to 12 hours each.

Propagation studies conducted at many locations around the world in microwave links experiencing the blackout mechanism (or in areas supporting such superrefractive layers) has revealed that if antenna heights establishing at least a k = 1 grazing clearance over an assumed 150-foot layer height are provided, blackout is avoided. This empirically-derived clearance standard has been used in the initial engineering, and subsequent correction, of numerous microwave paths with (thus far) total success. This does not presume that the layer will never rise above 150 feet, but rather that the receive antenna will remain within the radio horizon of the transmit antenna even during layering.

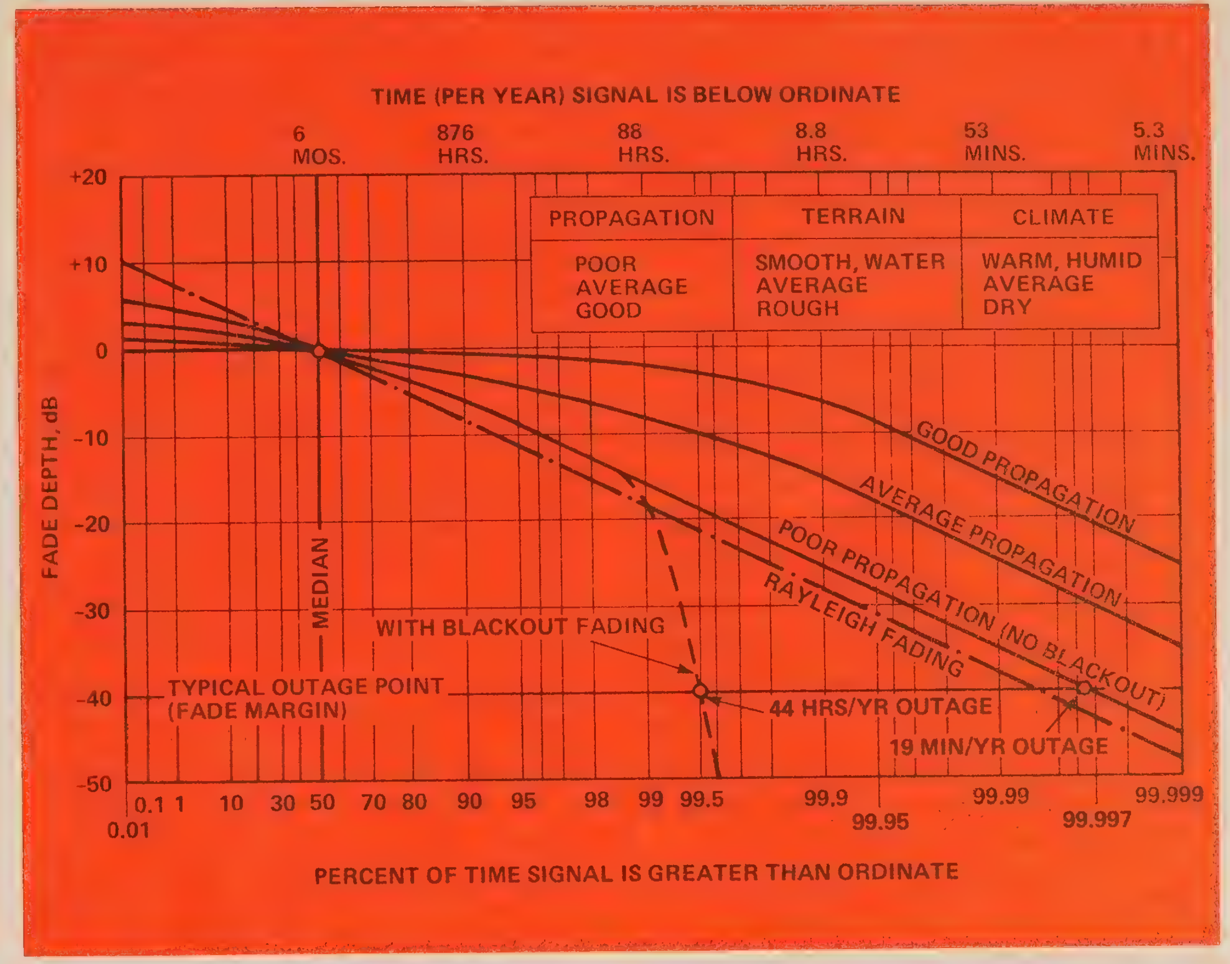


Figure 7. Cumulative probability distribution of fade depth with and without blackout fading on a 22-mile, 6.7-GHz link.

Many microwave paths traversing difficult propagation areas have been engineered with a grazing clearance over the terrain (or water) of k = 5/12. On paths which are longer than 25 miles, and with equal tower heights, this criterion dictates higher path clearance and antenna heights than the k = 1 grazing over a 150-foot atmospheric layer criterion, and is probably overconservative in any area. With path lengths between 15 and 25 miles, the k = 1 grazing over a 150-foot layer appears appropriate.

Blackout fades seldom occur in properly engineered paths shorter than 15 miles in length since the refractive index gradients are seldom so steep that the receive antenna appears beyond the radio horizon of the transmit antenna. But the climatology of any suspected area should be carefully investigated before selecting antenna heights. This is particularly important on paths configured in a high/low antenna arrangement (used to reduce interference fading) where the low antenna is near an area that may support a superrefractive layer.

Clearance Requirements

If a path is found susceptible to blackout fading after installation, the first step is to investigate the possibility of antenna decoupling, particularly if the parabolas or reflectors are large, the path is long, and the higher 6- to 13-GHz bands are used. If one or both antennas are initially oriented slightly down from the normal arrival angle of

the path, a superrefractive layer could easily refract the ray enough to miss the main antenna lobe, resulting in a blackout fade. The larger (or lower) of the two antennas could be reoriented vertically during the blackout period in searching for the path. A small 2- or 4-foot search receive test antenna could be used in lieu of reorienting the large path antennas to isolate possible decoupling. A smaller antenna might be eventually required in the path to prevent decoupling.

If the receive antenna is proved to be beyond the radio horizon during blackout, rather than being simply misoriented or decoupled, the second step is to search for the presence of a stable signal of normal level low on the tower, perhaps 10-30 feet AGL (above ground level), or even below normal optical line of sight. If such a path is found to exist during blackout, the receivers could be configured for space diversity, or the diversity antenna placed lower on the tower if already so configured. Special automatic or manual transmitter switching or antenna combining arrangements are required in hot-standby systems since blackout fades occur in both directions simultaneously.

If antenna orientation and signal search with a low antenna are both unsuccessful, the third step is to increase the path clearance to a minimum of k = 1 grazing over a 150-foot layer. An investigation should reveal the location of the layer and the antenna heights adjusted at that end. If the suspected layer is mid-path, both tower heights could be increased. Most towers may be extended somewhat, particularly if the antenna load is less than the original design, although it may be necessary to reduce antenna sizes, and receive signal levels, to permit the extension of a self-supporting tower. Guyed towers are often extended with little structural modification, perhaps by simply increasing guy diameters and incorporating double lacing in overstressed parts of the structure.

A nominal reduction in fade margin to correct a blackout situation, such as reducing antenna sizes or increasing waveguide loss, has little affect on the outage time due to this catastrophic fade. But preventing antenna decoupling or raising the path above the layer could have a significant influence on the path, hopefully eliminating any future occurrence of blackout.

If these steps are unsuccessful or are not economically or technically feasible, other major solutions to a blackout environment are: (1) adding a repeater to shorten the path length, thus bringing the receive antennas to within the radio horizon of the transmitter during blackout; (2) rerouting the path around the blackout area with existing stations, by changing a backbone repeater to a spur terminal, for example; or (3) abandoning one or more repeater or terminal locations and relocating them to areas that support more congenial climates. Complete abandonment of an entire microwave system in favor of cable or other communications media not susceptable to the blackout mechanism is, of course, the least desirable solution.

Climatology

Although ground-based observations of temperature and humidity are useful in revealing atmospheric layering to a knowledgeable observer, a more precise and meaningful method is by radiosonde or refractometry measurements. Radiosonde observations (RAOB) sample upper air temperature and humidity as a function of altitude by airborne instrumentation, and transmit this information to a ground station. Radiosonde measurements are

made twice a day at many weather stations; these measurements, however, are usually made in the morning and evening, before and after the critical 10 p.m. to 6 a.m. period when most blackout layers occur. Direct refractive index measurements can be made with a refractometer, a device that may be sent aloft by balloon or suspended under the wing of an aircraft descending into the suspected blackout area.

Radiosonde data, when plotted on adiabatic charts, clearly reveal trapping layers. The U.S. Weather Bureau recently went to code for efficient storage of these data, reducible to chart form by referring to the Manual for Radiosonde Code (WBAN), available from the Government Printing Office, Washington, D.C. The actual radiosonde (and ground-based) data for any domestic area under investigation are available from the National Weather Records Center, Asheville, NC.

Environmental Indicators

There are several environmental indicators which may give a clue to the formation of atmospheric layering conducive to blackout fading. One of these is changes in water temperature with the ambient temperature; this could occur in shallow bodies of water such as swamps, irrigated fields (seasonal), and areas close to shorelines. Layering of smokestack smoke at low altitudes, or the presence of low-lying mist (steam fog), are also indications of possible atmospheric layering.

The formation of a visible haze layer on an otherwise clear day is another indication of possible layering. For example, it is occasionally possible to see, from atop a microwave tower, distant mountains above an atmospheric discontinuity, but to have the mountains disappear if the observer is within or below the layer. An escarpment in New Mexico was named Fade Away Ridge by early settlers since it would occasionally disappear from view; microwave paths in this area, north of Carlsbad, are highly vulnerable to blackout fades severe enough to have required the relocation of several links and the abandonment of at least one microwave station.

The occurrence of blackout fading in some particular geographical locations is seemingly inevitable during times of the year when deciding factors such as atmospheric density and nature of terrain are right for the occasion. However, with an intimate knowledge of the climatic characteristics of a suspected area, the microwave engineer has an excellent chance of avoiding the devastating effects of anomalous propagation.

BIBLIOGRAPHY

1. Bean, B.R. and E.J. Dutton. *Radio Meteorology.* New York: Dover Publications, Inc., 1966.

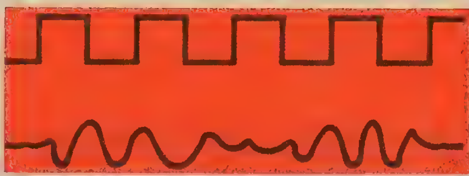
2. Laine, R.U. *Determining Microwave Antenna Heights for Coastal Areas (Monograph 197).* San Carlos, California: GTE Lenkurt Inc., 1964.

3. Samson, C.A., et al. *Weather Effects on Approach and Landing Systems (ESSA/ITS Report No. FAA-RD-70-47),* July 1970.

4. White, R.F. *Engineering Considerations for Microwave Communications Systems.* 2nd Ed. San Carlos, California: GTE Lenkurt Inc., 1972.

SECTION IV
DATA TRANSMISSION

GTE LENKURT
DEMODULATOR
OCTOBER 1973



Is it digital, digitized-analog, or analog? These distinctions need to be understood when designing and using telecommunications equipment.

An abacus with its two-position beads represents a digital system; while a slide rule, with its infinitely variable slide positions, is considered an analog device. The non-digital answer that results from using a slide rule becomes quantized and digitized as soon as the slide rule product is conceived as a discrete numerical value.

Digital data is defined simply as information expressed by a set of discrete values or symbols. These can be based on various systems – if in a computer, on the binary numbering system with "ones" and "zeros"; if a printed page, a set of 26 letters and 10 numbers. Analog data, by contrast, can theoretically assume an infinite number of values within a prescribed range. Analog examples include the loudness of a talker's voice, the varying velocity of the wind, and the result of a slide-rule calculation. However, all measurements of analog information imply approximations. For example, if a result from a slide rule is read as "628" while a more correct result really is 627.95, the eye has made an approximation because it cannot detect the difference. Rounding off to the nearest number in a set of discrete values is called "quantizing" and effectively converts analog data into digital data. In a digital system, these approximations are built into the initial coding scheme. The greater the number of possible codes (or quantizing levels), the more accurate the approximation.

Furthermore, while a printed word is digital, the same word when spoken, with its frequency and amplitude variations, becomes analog, and will represent a different analog signal when spoken by two different people (although the written information being conveyed through speech will generally be the same). This analog signal can be electronically sampled and then quantized into a set of discrete amplitudes for transmission over a digital communications system. At the other end of the communications link, this digitized signal can be converted back into a speech signal. Although still digital in the sense that only a discrete number of choices are available for reconstruction of the speech signal, it closely resembles the original analog voice, especially at low speech volumes. If further desired, a speech recognition system can be employed to convert this reconstructed speech signal into a written or printed word. If errors have been kept to the required minimum, this final reconstructed word should convey the same information as the original.

Another way to transmit this printed page would be to use an all-digital transmission system such as telegraph. But, if analog voice signals of transmission or reception are desired along the

way, it may be necessary to use both digital and analog means of information transmittal.

Digital

Digital communication involves the transmission of discrete coded symbols over a transmission channel which has a potential for a high noise level. A digital receiver must be capable of correlating the received signal with this finite set of digital values or symbols, and if necessary, converting to an analog signal. Generally, if the S/N (signal-to-noise) ratio is below a certain threshold, the reproduced digital signal will be free of errors. With an analog signal, the noise, regardless of its level, will distort the signal proportionately to the S/N ratio. If there are any repeaters in the transmission system, the digital signal will be regenerated, free of any noise, at each repeater; while the noise associated with an analog signal accumulates at each repeater.

In digital machines, information is expressed in digital form and the digits are represented by the states of certain physical parts in the machine which can assume one of a finite set of possible states such as the "on" or "off" states of switches. Typical digital machines are the abacus, ordinary desk calculators, telegraph, and typewriters. The precision of a binary digital machine increases exponentially with the number of digits or possible states (see Figure 1), and hence the machine's complexity. Small noise levels in a digital system have no detrimental effect and do not accumulate unless they are significant enough to change the signal to another state.

Binary digital machines are ideal for transmission systems that can tolerate very few errors, and they have the design advantage that there are a variety of two-state components available. Since there are only two states that need to be distinguished in a binary-coded message, the noise level in the transmission system can be quite high before it causes errors in reception of the message. The value of a physical form expressed as a binary digital word is determined by the presence of a one or a zero in each of the digits of the word. Therefore, expressed in terms of voltage levels, a binary digital system requires only two discrete voltage levels. The greater the difference between these two levels, the greater can be the noise in the system before any errors are introduced.

Analog

An analog signal is simply a signal that is continuously variable within a range of values. The signals from such equipment as telephone, radio, and television are generally analog in nature. These transmission systems have traditionally been analog because digital circuit technology has only become practical within the last 25 years.

It is interesting to note that Samuel Morse's digital telegraph (1832), which preceded Bell's analog telephone system (1874) by more than 40 years, was the inspiration for Bell's invention. And, digital transmission is now gaining importance for telephone systems.

In an analog system, data is represented by measurements on a continuous scale, so that the accuracy of the machine is determined by the accuracy of the scale. An analog signal is usually represented as a voltage, phase, current, or frequency, the magnitude of which gives the value of some physical quantity like temperature or time. Analog signals are also represented by such things as color shades or saturations, height of an indicator, angle of a plumb-bob, or the deflection of a compass needle.

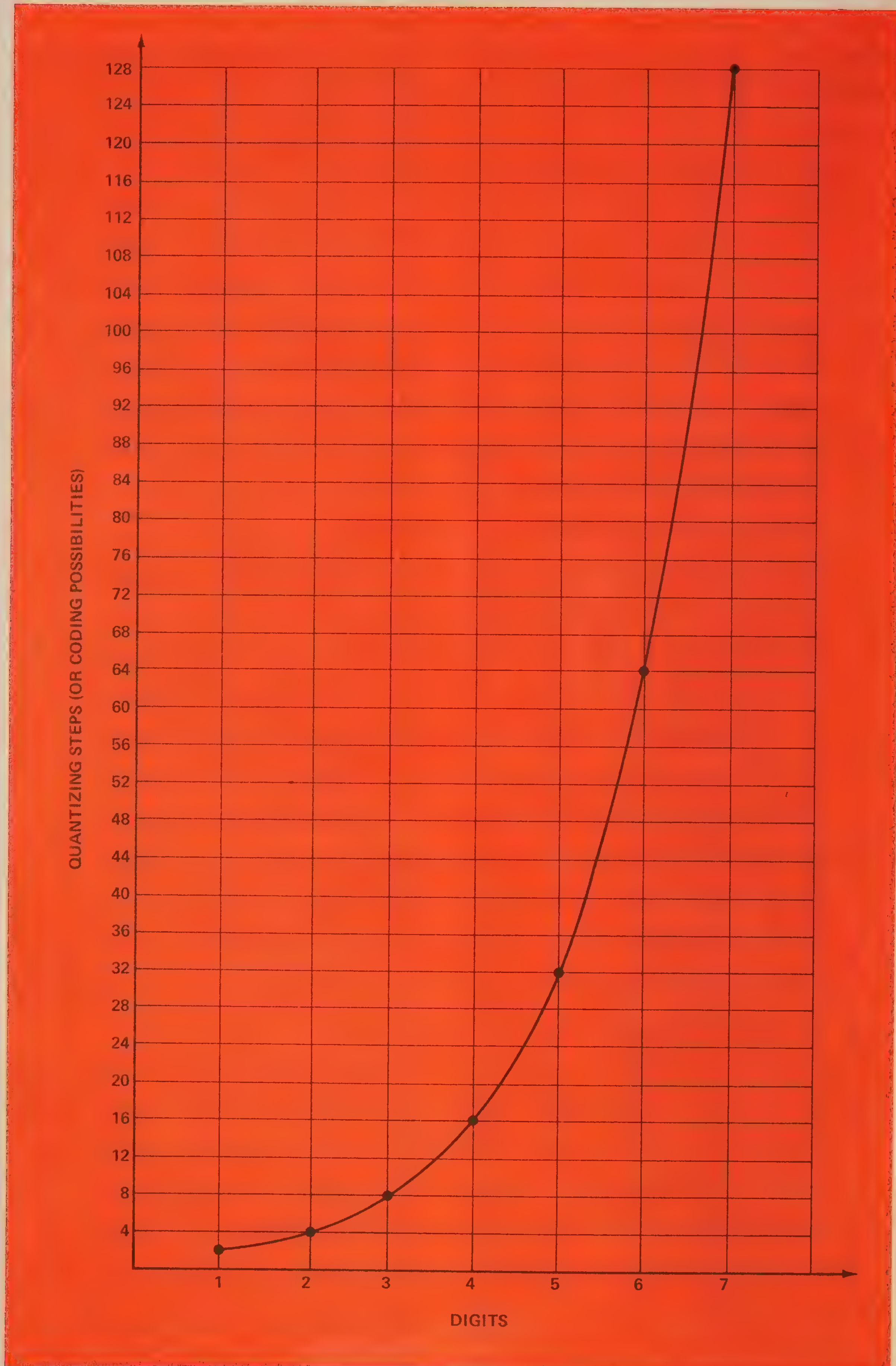


Figure 1. The accuracy of binary digital coding schemes increases exponentially with the number of digits available for coding.

In analog machines, the outputs represented are proportional (that is, analogous) to some physical quantity capable of continuous variation. Typical examples are the slide rule, thermometer, and D'Arsonval meter movement. An increase in precision generally requires a propotional increase in the range of physical variables used to represent the numbers. Furthermore, small errors tend to accumulate, and cannot be eliminated. Consequently, in an analog system, to decrease noise susceptibility by 2 to 1 requires roughly a 2 to 1 increase in the swing of the frequency or the amplitude, which causes an increase in the signal bandwidth by an even larger factor, because more higher-order sidebands become significant.

Digital on Analog

Eyes and ears are analog receptors of information. These receptors can also be used for digital information. Eyes can recognize discrete states and distinguish between these states. They are also receptors for infinitely variable color shades and hues. Ears can distinguish between individual tones and at the same time fully appreciate continuously varying tones such as found in musical compositions.

The same is true of analog telecommunications equipment. As long as the digital signal being transmitted is within the spectrum – frequency or amplitude – of the analog system, there is no difficulty in transmitting the digital signal on the analog channel.

The reverse is not true without some additional steps. An analog signal cannot be transmitted and received on digital equipment unless the analog signal is first sampled, quantized, and then coded according to some predetermined standards. Then at the receiving end, this digitized signal is restored to its analog state.

Data modems are generally used when transmitting a digital signal over an analog channel. The modem is responsible for making sure the digital signal is in the proper part of the spectrum for transmission, and also for making sure that both ends of the transmission system are properly synchronized so that the data received corresponds with the data sent. What happens to digital information when it goes through a data modem is analogous to what happens to an analog signal when it is placed in the proper frequency spectrum and sent over a multiplex channel. In fact, many data modems are also multiplexers so that the function of a data modem and a voice multiplexer can be considered the same. If the modem is not a multiplexer, the output signal from the modem can then be put through a multiplexer the same as an analog signal.

The function of a data modem is not to perform analog-to-digital conversions or digital-to-analog conversions. If a digital signal is put into a modem, a digital signal will be the output of the modem. The same is true of analog systems – analog in, analog out.

Therefore, it should be remembered that when an analog channel is used for transmission of digital data by the use of a modem, the digital data is not really "made into analog", but is rather remade into a shape and format *looking much like* analog (see Figure 2). The purpose is to utilize a system that was originally designed for analog transmission (like the telephone network) for digital transmission. Since, at the receiving end, the intent is to distinguish between a finite number of discrete states for decoding of the digital signal, the signal is, in effect, still digital when transmitted over the system. The fact that a train of

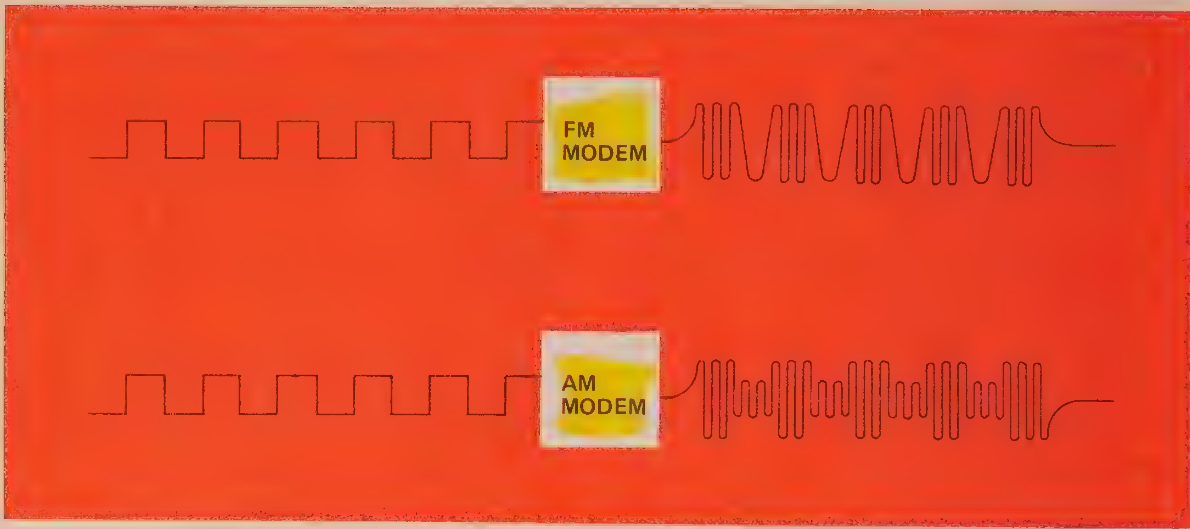


Figure 2. A data modem transforms a digital signal into an "analog" looking signal that is made up of two discrete frequencies or amplitudes.

"square" looking pulses is made into a more rounded shape or a set of frequencies, phases, or amplitudes does *not* make it into analog – it just looks more like analog. It is still a digital signal.

Differentiation Between Analog and Digital

A circuit designer with the requirement to transmit the number 100 has a choice to make between digital and analog. One way to transmit this information is to send a voltage of 100 volts and measure the voltage at the other end with a voltmeter. This would be analog transmission. This same number 100 can also be sent by using a seven-bit binary code to represent 100 (1-1-0-0-1-0-0). By going from a seven-bit binary code to an eight-bit binary code the accuracy of the system is doubled, but the bandwidth need only be increased by 15 percent since only one additional digit is needed. Figure 3 shows the increased accuracy versus bandwidth for a binary digital system.

By recognizing the differences between analog and digital transmission the designer can make the best possible use of the commodities that he has – money, bandwidth, and time – and best satisfy his needs for accuracy. There are definite trade-offs that can be made between digital and analog transmission when considering bandwidth and data rate. Since spectrum conservation is highly important these days, it is imperative to design transmission equipment to make the most efficient use of the available spectrum.

Just because a system is digital doesn't mean that the available spectrum has been used in the most efficient manner. An example of this is the difference between a rotary dial telephone and a tone dial telephone. The rotary dial is a digital signaling scheme and for each digit that is dialed it is necessary to step off the correct number – for the digit seven, step switches are required to step seven times. This is a relatively slow process compared to tone dialing. Tone dialing is a digital signaling scheme in which five different tones (single frequency signals), working in pairs, represent the ten digits used in telephone dialing.

Interoffice signaling has used a similar two-out-of-five digital scheme for some time due to the speed of digital

NO. OF DIGITS	NO. OF QUANTIZING STEPS	APPROXIMATE BANDWIDTH INCREASE (%)
1	2	100
2	4	50
3	8	34
4	16	25
5	32	20
6	64	17
7	128	15
8	256	13
9	512	11
10	1024	10

Figure 3. Maintaining the same noise threshold for a binary digital system, quantizing accuracy can be increased substantially with a diminishing bandwidth, or data rate, increase.

signaling. In the North American telephone industry, the saving of one second of equipment set-up time could result in a multi-million dollar saving.

Common Misconception

In the telephone industry, it is commonly thought that PCM (Pulse Code Modulation) is the only form of digital transmission. This is a misconception since a digital signal can be sent over a frequency modulated or amplitude modulated channel after passing through a modem to make the digital signal "look" analog so that it is compatible with the transmission channel.

Finally, it is often thought that if the signal is digital, the transmission media is digital. But, regardless, of the signal form – digital or analog – the transmission media, whether wire, cable, or radio frequency, is simply a highway over which the signal travels. If the signal is regenerated at repeater points, the channel could be classified as digital; if the repeater simply amplifies, it is analog. But the transmission media itself doesn't care.

BIBLIOGRAPHY

Communications-Electronics Terminology Handbook. Washington, D.C.: Public Affairs Press, 1965.

Cooke, Nelson M. and John Markus. *Electronics and Nucleonics Dictionary.* New York: McGraw-Hill Book Company, Inc., 1960.

Hoeschele, David F., Jr. *Analog-to-Digital/Digital-to-Analog Conversion Techniques.* New York: John Wiley & Sons, Inc., 1968.

Oliver, B. M. and J. R. Pierce. "The Philosophy of PCM," *The Proceedings of the I.R.E.*, (November 1948), 1324-1331.

Rabiner, Lawrence R.; James W. Cooley; Howard D. Helms; Leland B. Jackson; James F. Kaiser; Charles M. Rader; Ronald W. Schafer; Kenneth Steiglitz; and Clifford J. Weinstein. "Terminology in Digital Signal Processing," IEEE Transactions Audio Electroacoustical, Vol. AU-20, (December 1972), 322-257.

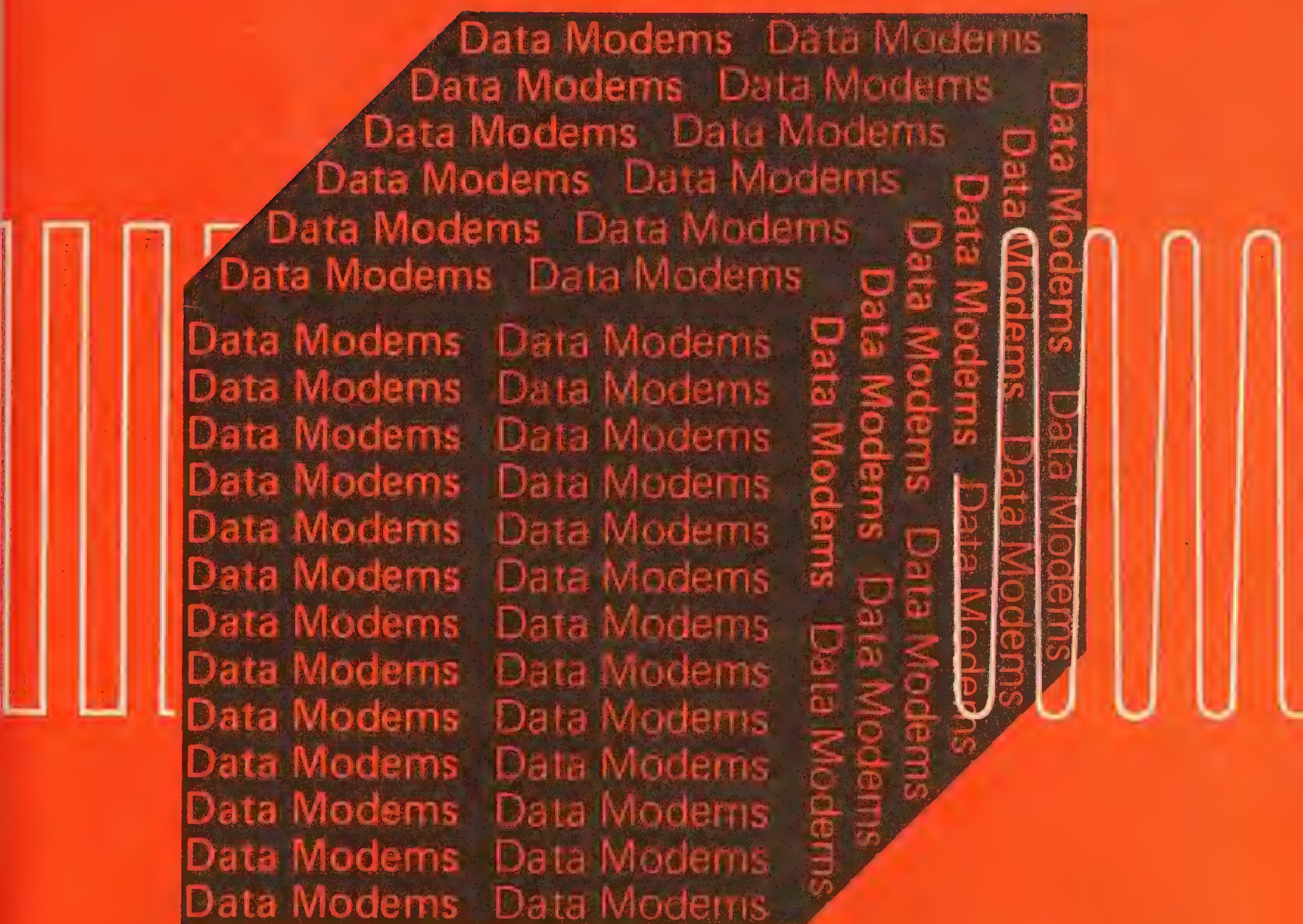
Ristenbatt, Marlin P. "Alternatives in Digital Communications," *Proceedings of the IEEE*, Vol. 61, No. 6, (June 1973), 703-721.

Wiener, Norbet. *Cybernetics or Control and Communications in the Animal and the Machine.* Cambridge, Massachusetts: The M.I.T. Press and John Wiley & Sons, Inc., 1961.

GTE LENKURT

DEMODULATOR

OCTOBER 1971



Modems . . . those unglamorous but vital "black boxes" that form the interfaces between digital machines and the communications network.

Modems have been referred to as data sets, line adapters, modulators, or subsets. Regardless of the name used, the purpose of each of these black boxes is to convert digital pulses into analog signals, such as audible tones, suitable for transmission over the telephone network.

If two machines – such as computers, data terminals, or facsimile machines – are communicating via the telephone network, it is necessary to have a modem at each end of the line to act as the interface between the machine and the communications line. These black boxes must be capable of modulating and then demodulating the signals – hence, the contraction modem from *mo*dulate and *dem*odulate.

A pair of modems are considered "transparent" since the signals into the first (the input) are identical to the demodulated signals (the output) from the second modem. Figure 1 illustrates the position of the modems in a data link and the signals into and out of each element of the link.

The modem and the communications line can be connected directly (hardwire) or indirectly (acoustic or inductive coupling). Acoustically coupled modems are portable since they can be used with any available telephone. With acoustic coupling, the dc data signals are converted to audible sounds which are picked up by the microphone (or transmitter) in an ordinary telephone handset. The audible signal is converted to electrical signals, and transmitted over the telephone network. The process is reversed at the receiving end.

Inductive coupling, like acoustic coupling, requires no direct connection. With inductive coupling a data signal passes to the telephone through an electromagnetic field by way of a hybrid coil.

Acoustic/inductive couplers generally do not operate as reliably as direct electrically connected modems, because they involve an extra conversion step (for example, digital to audible to electrical) where noise and distortions may be introduced. For this reason acoustic/inductive modems are presently limited to transmission speeds below 1200 bps (bits per second).

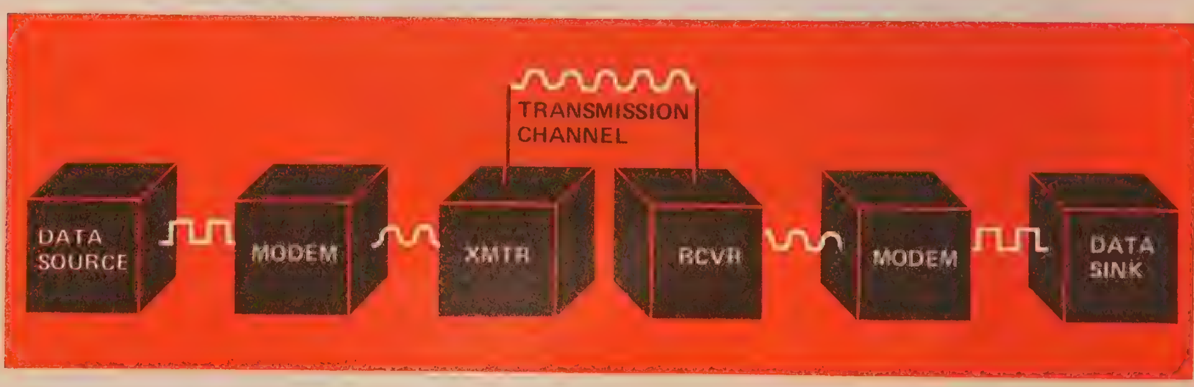


Figure 1. A data link requires a modem on each end of the transmission channel.

A direct hook-up to the communications channel, therefore, is preferable, since it is less error-prone and not limited to low speed transmission.

Asynchronous or Synchronous

Having connected the digital machine to the communications line it is necessary to coordinate the data received with the data sent. This coordination or synchronization can be accomplished in two ways. If start and stop bits are used to "frame" each character, the transmission is asynchronous. Synchronous modems require the use of clocking devices which lock the transmitted signal of the modem and the terminal device together at a fixed transmission rate.

High-speed data generally uses synchronous transmission since for identical data coding levels and transmission bit speeds, a higher data speed can be achieved. Asynchronous transmission requires the use of two or three start- and stop-bits for each character depending upon the type of machine generating the digital signal. Consequently, if an eight-bit code is being used, asynchronous transmission requires 10 or 11 bits per character and synchronous requires only 8. Synchronous modems can therefore transmit at least 25% more characters than asynchronous modems at the same bit speed (see Figure 2).

Although synchronous transmission is efficient, the clocking mechanism requires added circuitry which makes the equipment more costly than asynchronous modems for the same speed.

Asynchronous modems have a specified maximum transmission speed, but they can be used to transmit data at any speed up to this maximum. Asynchronous modems are used for low- and medium-speed transmission up to approximately 1800 bps.

High-speed modems, on the other hand, are intended for synchronous operation at a fixed transmission rate. The transmission speed of a synchronous modem is established by the clocking source which is generally a crystal oscillator. If a synchronous modem has more than one speed, the speeds are generally multiples of the oscillator frequency.

Parallel or Serial

Another way to classify data modems is according to the type of bit stream used – parallel or serial. Figure 3 illustrates the difference between serial and parallel bit streams. Serial bit streams are most commonly used since the digital information can be modulated as it comes from the digital machine. As long as sufficient bandwidth is available for transmission without degradation, a serial bit stream may be used.

If, however, transmission is to take place over bandwidths which do not have uniform transmission characteristics, a serial bit stream can be converted to a parallel bit stream. At the receiver, a parallel-to-serial conversion

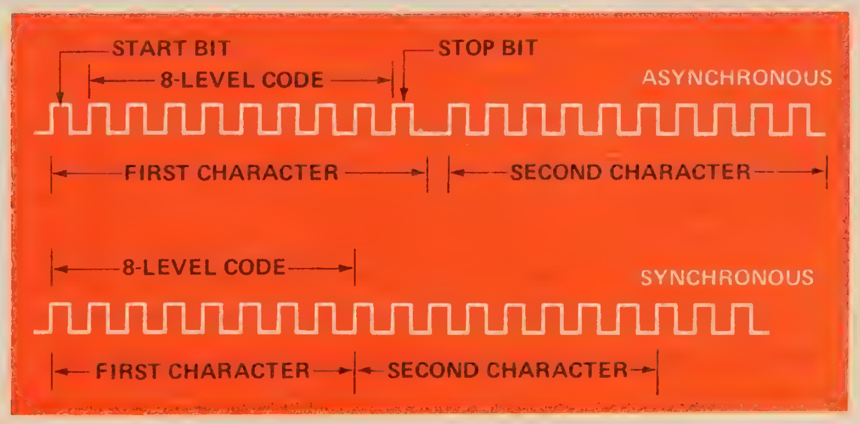


Figure 2. Each character is at least 25% longer with an asynchronous system than with a synchronous system.

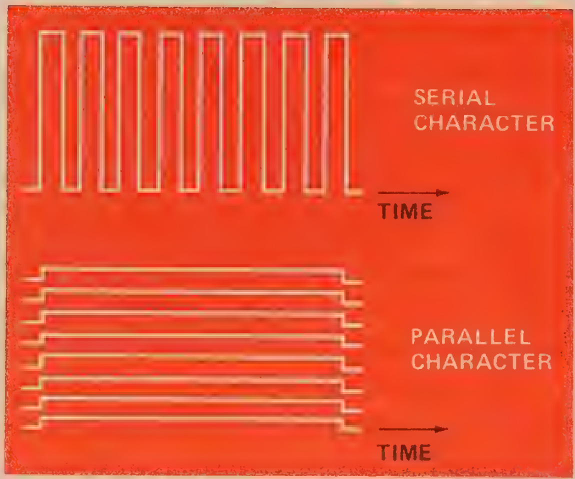


Figure 3. With parallel transmission, longer bits are used to transmit the same amount of data in a given period.

takes place. This technique is often used to transmit data at 4800 bps and higher over a voice channel.

Parallel channels with their longer symbols provide better correlation of fade and phase factors and multipath delay distortion in the propagation medium — radio or cable. (See September, 1970, *Demodulator* for discussion of transmission impairments.) However, the complex circuitry for parallel transmission makes parallel modems more costly. They are also less efficient since bandwidth is used for flanking of the bandpass filters in each channel. For these reasons serial modems have been accepted as an industry standard.

Modulation

Modems can also be classified according to the analog signal generated in the D/A (digital to analog) conversion. These analog signals may be amplitude, frequency, and phase modulated. Four types of modulation are used extensively in digital data transmission: amplitude modulation (AM), frequency modulation (FM), phase modulation (PM), or AM combined with either FM or PM.

With an AM modem, the sinusoidal carrier wave is varied in amplitude to correspond to the digital information being transmitted. Upper and lower sidebands equal to the carrier plus the modulating signal and the carrier minus the modulating signal, respectively, occupy a total bandwidth of twice the modulation rate (see Figure 4). The entire double-sideband AM signal, a single-sideband AM signal, or a vestigial-sideband AM signal can be transmitted depending upon how the signal is processed at the receiving end.

In single-sideband only, one sideband is transmitted with or without the carrier, and the required transmission bandwidth is only half that required by double-sideband AM. But, if it is necessary to transmit a dc component for signal processing, vestigial-sideband AM must be used — transmitting the wanted sideband, part of the carrier, and the low frequency end of the unwanted sideband.

Single-sideband AM gives the best bandwidth economy, but not the best equipment economy. The filtering necessary for single-sideband AM is difficult to achieve; consequently, the technique is used primarily for high-speed data transmission over a band-limited channel where the advantages outweigh the disadvantages.

Vestigial-sideband AM systems require a bandwidth approximately 1.3 times that required for single-sideband AM systems, and the technique is typically used in data modems operating at speeds of up to 7200 bps over voice-grade lines.

In AM transmission the amplitude of the carrier is varied but with FM transmission the carrier frequency varies proportionally to the instantaneous value of the modulating signal — the data bit stream. When transmitting binary data, the frequency of the transmitted wave shifts between two discrete values (determined by the channel bandwidth), one representing binary one and the other, binary zero. This is a double-sideband system called frequency shift keying (FSK) and re-

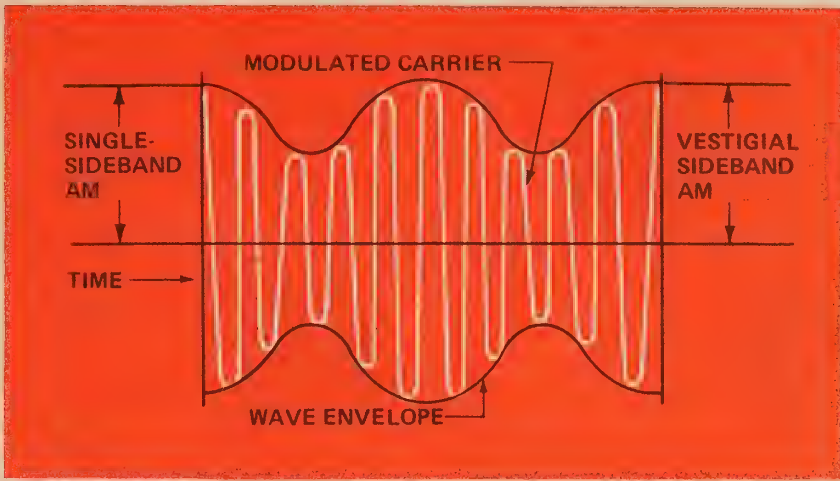


Figure 4. The wave envelope is the same as the modulating data signal. The bandwidth of the double-sideband AM signal is twice the bandwidth of the modulating signal.

quires approximately the same bandwidth as double-sideband AM.

For phase modulation the phase of the transmitted carrier varies proportionally to the instantaneous value of the modulating signal. For binary data transmission the phase is shifted 180° for each transition between one and zero, or zero and one. Phase shift keying (PSK) is extensively used for synchronous high-speed data transmission systems up to 2400 bps. However to transmit at 2400 bps and above, it is necessary to resort to a four-phase system utilizing 90° shifts. PSK, like FSK, is a double-sideband system.

In a special form of FSK called duobinary, FSK modulation is used in conjunction with duobinary coding which uses a three-level code to represent the binary data. The duobinary coding technique, developed at GTE Lenkurt, is used in the GTE Lenkurt 26C data modem. Assuming a constant bandwidth data channel, duobinary FSK transmits at twice the speed of FSK. Figure 5 shows the difference between FSK and duobinary FSK.

Factors to consider when selecting a modulation scheme are complexity of electronic circuits, required bandwidth, quality of transmission channel, signal-to-noise ratio, tolerance to delay distortions, tolerance to amplitude changes, tolerance to jitter, and reliability. Each system has some advantages relative to the other systems.

Compared to Voice Band

Another way to classify modems is by their required transmission bandwidth. Modems can be divided into three categories: sub-voice, voice, and greater-than-voice band (or wideband). The chart in Figure 6 relates bandwidth to transmission speed, and illustrates that transmission speed is generally proportional to bandwidth.

Sub-voice band modems use only a fraction of the 4-kHz voice band for each data channel. These modems generally serve slow digital devices with speeds of up to 600 bps. Frequency division (FDM) or time division (TDM)

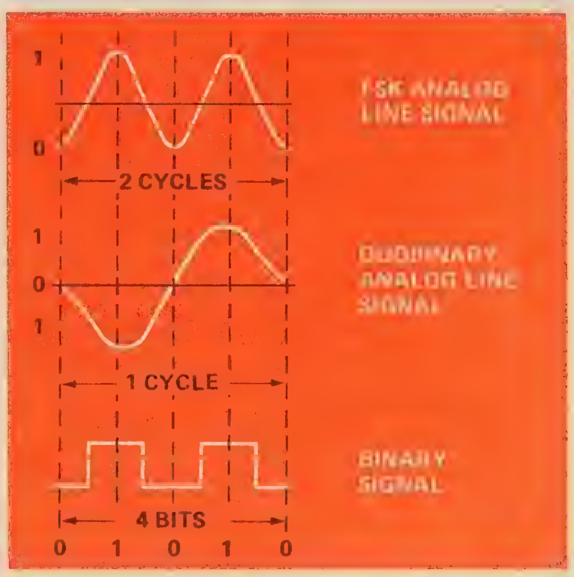


Figure 5. Duobinary FSK transmission is twice as fast as FSK transmission showing four binary bits compressed into one cycle of the line signal.

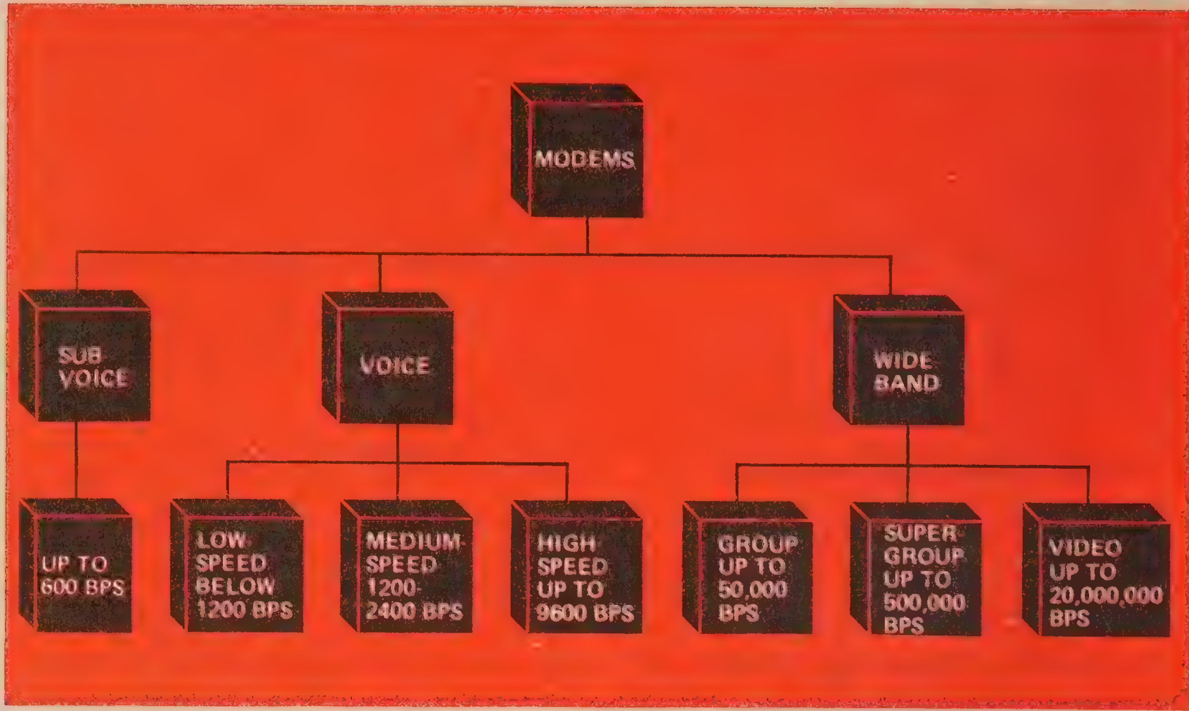


Figure 6. Bandwidth and transmission speed are generally related as shown in this chart – the wider the bandwidth, the higher the speed.

multiplex techniques are used to fill the voice band. With FDM, a multiplexer is not needed because the modem conditions the data signal for transmission in its proper frequency slot on the telephone line – the signal is in essence frequency modulated and translated. GTE Lenkurt's 25C data modem performs such a multiplexing function. But, with TDM, a multiplexer combines the digital signals in time. This new high-speed, serial bit stream then goes to a modem for digital to analog conversion and transmission over the telephone circuit. Figure 7 illustrates the difference between FDM and TDM systems.

FDM and TDM are equally suitable for voice grade channels. However, TDM is more efficient in bandwidth utilization; therefore, more channels can be multiplexed on a single channel. Conversely, FDM is best suited where few circuits are dropped off and picked up at scattered points, and where the greater reliability of individual channel modems is desired.

Since voice band modems range in speed from as low as 300 up to 9600 bps or higher, they are not defined as much by speed as by the facility and means of transmission. Voice band data is the most efficient means of utilizing the telephone network.

Satisfactory modem performance at 3600 bps and above generally requires complex equalization circuitry to precondition the signal for non-linear or varying line parameters – such as delay distortion and attenuation. Some high-speed modems, generally used over leased lines with relatively constant characteristics, use manually adjusted equalizers. Other high-speed modems use automatic or adaptive equalizers to continually adjust to the line characteristics.

Medium-speed voice band modems operating in the 1200- to 2400-bps range have been in general use for over ten years and have achieved a degree of reliability and low-error performance adequate for most data transmission applications. Most of these medium-speed modems tolerate or pre-condition the signal for minor changes in telephone line characteristics without error or interruption of

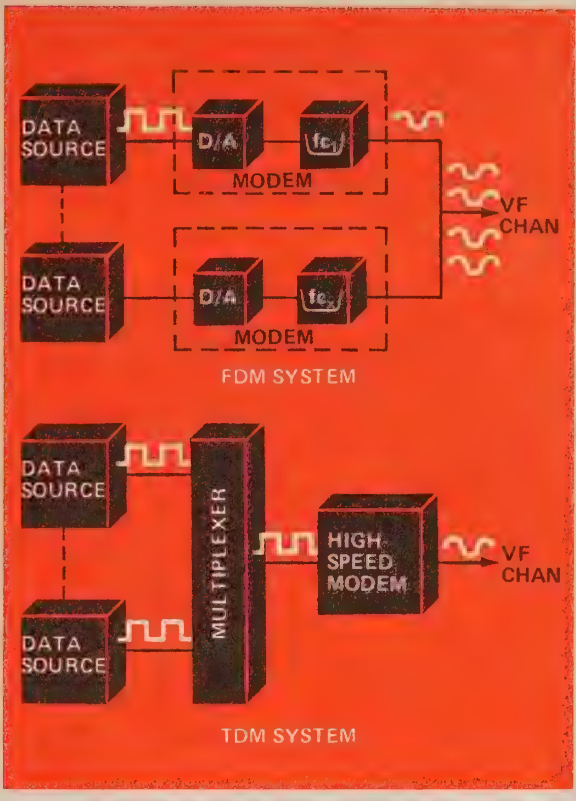


Figure 7. A time division multiplex system requires the use of a multiplexer and a modem.

transmission. Medium- to high-speed modems at 4800, 7200, and 9600 bps are finding wider use as transmission speeds continue to increase.

The telephone network is designed such that when the signal bandwidth exceeds one voice channel, the next transmission channel has a group bandwidth or supergroup bandwidth – equivalent to 12 or 60 voice channels. Greater-than-voice band or wideband modems operate at speeds from 18,750 to 500,000 bps. Top speed is presently limited by the expense of leasing wider bandwidths and the limited need to move much larger volumes of data at these rates.

Wideband modems are not true modems since they do not contain a modulator or demodulator, but they do condition a digital signal for transmission over the telephone network. In wideband data sets, the digital signal is first put through a scrambler which inverts every other pulse to eliminate sustained intervals of ones or zeros that might create an undesirable dc component in the line. Next the signal is filtered to remove low- and high-frequency components. The result is an ac signal which, for 50,000 bps data, has a 25-kHz fundamental frequency – since two bits complete a cycle, the fundamental frequency is half the bit rate. There is no need to translate the wideband signal in frequency, as there is for sub-voice and voice band modems.

This ac signal readily passes through transformer-coupled circuits and over non-loaded physical cable pairs with reasonable equalization and amplification. The signal can also be fit into a twelve-channel bandwidth for analog exchange and trunk carrier systems.

High-speed modems may be frequency division multiplexed to put a number of them in parallel on a single wideband circuit.

The great advantage of digital transmission via wideband systems is that high data transmission rates may be obtained while keeping the data stream in serial form. In the multiplexing equipment it is not necessary to come down to the nominal voice channels (4-kHz), but the serial data streams may be modulated on a group bandwidth (48-kHz) with a data rate capability of 50,000 bps, or a supergroup bandwidth (240-kHz) at up to 500,000 bps.

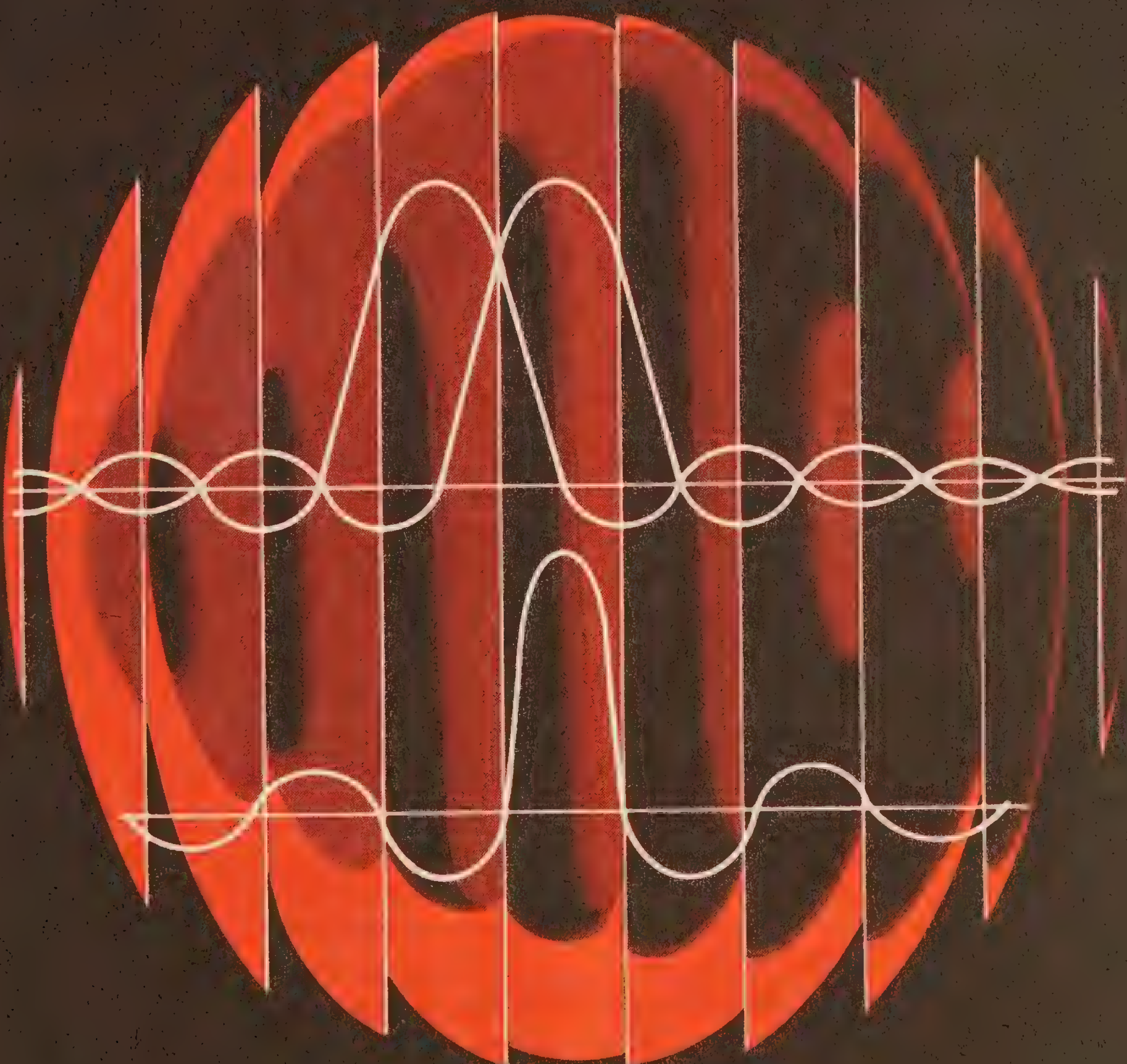
All-Digital Network

All-digital transmission networks which would not require data modems are being designed, developed, and tested, but it will be a long time before the sub-voice and voice band data systems requiring modems are eliminated from the telephone network – if they ever are.

GTE LENKURT

DEMODULATOR

AUGUST 1974



data transmission: principles and problems

> **"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point." C.E. Shannon**

In moving from primitive to modern, man has consistently sought for ways to make tasks less difficult or tedious; this search has almost invariably resulted in the invention of machines. Communications "machines" have run the gamut from papyrus rolls and rune staffs to printed books and computer print-outs.

Data Language

Communication implies a language or symbology; almost without exception, the language employed in communication between machines consists of electrical signals, with the information conveyed in a digital form.

Digital symbology is based on the premise that all information can be conveyed in a two-state code. Samuel Morse applied this principle to the development of his code, which uses long and short transmission times as coding to represent the letters of the alphabet, numbers, and various symbols.

The two-state code concept is derived by considering information to be present only where there is doubt; that is, where a choice, selection, or discrimination is required. In other words, a message is considered to have the potential to communicate something, with the information communicated dependent upon the selection of one message from a set of messages. The simplest choice is between two equally possible messages, which may be yes-or-no, on-or-off, A-or-B, 0-or-1, or any other two-state condition. As an example, figure 1 presents a group of eight messages, 1 through 8. Assuming that one has been selected by a hypothetical source and signaled to a receiver, and that all of the messages are equally likely to have been selected, the question is, how much information must be conveyed for the receiver to identify the proper message. Choice 1 asks the question, "Is it in the first half of the group, yes or no?" (yes = 1, no = 0). Having thus eliminated half of the possibilities, choice 2 asks, "Is it in the first half of this half?" Similarly, choice 3 determines in which half of that half the selected message lies. Should message 6 be selected, it could be located by the following responses: no (0) to choice 1, yes (1) to choice 2, and no (0) to choice 3. In this way, three simple yes-or-no (1-or-0) choices have served to identify one unique message out of a set of eight.

Since the two possible messages correspond to the two symbols in the binary number system, the information content of a message based on two-state coding is measured in units called binary digits, or bits. Expressed in bits, information content can be determined by the formula:

$$H = \log_2 m$$

where

H = the number of bits of information

m = the number of likely choices

If m is 8 — as it is in the example of figure 1, where eight possible messages are presented — the formula shows

CHOICE 1	CHOICE 2	CHOICE 3	MESSAGE
1	1	1	1
1	1	0	2
1	0	1	3
1	0	0	4
0	1	1	5
0	1	0	6
0	0	1	7
0	0	0	8

Figure 1. Amount of information required to identify a unique message from a set of eight equal possibilities.

that $\log_2 8$, or 3, bits of information are required to isolate a given message, which is what was found in the example. Finding one word in a dictionary of a quarter of a million words would require at most 18 bits of information, since 2^{18} is greater than 250,000.

A message consisting of one simple electrical pulse has the informational value of one bit, because the presence or absence of the pulse permits the receiver to select the correct message from a set of two. Mechanical and electrical devices, which can be made to change states very rapidly — as from conduction to non-conduction — are ideally suited to application of this principle.

Communication Systems

A general communication system consists of five elements:

1. An information source, which selects a message from a set of possible messages.
2. A transmitter, which transforms the selected message into a signal that can be sent over the communication channel.
3. A communication channel. This is the physical medium over which the signal is transmitted, and may be air, as in oral speech, a pair of wires, a coaxial cable, a beam of light, or simply space with the signal in the form of an electromagnetic wave.
4. A receiver, which performs the inverse operation of the transmitter to reproduce the message from the signal.
5. A destination, which is the person or machine for which the message is intended.

In digital communication systems, the information source and destination are most commonly computers, tape terminals, or business machines. Data from the source are typically represented by binary signals (messages) at the input to the transmitter, which is generally a modulation or encoding device, whose purpose is to transform the binary signal into another digital signal that requires less channel bandwidth. In data applications, it is normal to refer to two channels: the modulation channel and the coding channel. While both are physically part of the same communication system, the modulation element is considered to be analog and to consist of the transmission medium and the physical transducers needed to couple the signal to the medium. The coding element is concerned with the digital data; it is a digital link which not only includes the modulation channel, but is also extended to incorporate the encoding and decoding devices.

Bandwidth Conservation

The need to use less channel bandwidth is due to the band limitations placed on transmission media. If there were no limits, rectangular binary pulses could be sent directly through a communication channel; in reality, however, bandwidth is limited and its conservation — through maximum utilization — is of extreme importance.

It was recognized as early as 1924 that each pulse in a data stream can be made to convey additional informa-

tion if one or more of its characteristics is varied in a prescribed manner. In this way, not only the presence or absence of a pulse, but its phase, frequency, and amplitude levels as well, can be used to transmit information. The use of multilevel systems represents a significant improvement in bandwidth utilization over binary transmission since, for a given bandwidth, more information can be conveyed by making each pulse represent a group of binary digits, rather than a single digit as in binary systems.

However, increasing the number of levels in a signal while maintaining constant power makes that signal more sensitive to noise, and more inclined to produce errors; in addition, complex equipment is required to generate the multilevel pulses. Because of this, alternatives to multilevel systems have been sought which use fewer levels for the same bandwidth and bit speed. One method, developed by GTE Lenkurt, provides a two-to-one bandwidth compression relative to binary signaling; that is, the duobinary technique, as it is called, affords twice the transmission speed in bits per second as a binary system of the same fixed bandwidth. To achieve this with multilevel techniques, four-level coding would be required; a duobinary system does it with three levels, and the equipment needed is comparable in complexity to that of a straight binary system.

In the duobinary system, the encoding and decoding of a signal involves dealing with the algebraic sum, at the sample instant, of the value of the present digit and of interference from other digits. To accomplish this, a binary input signal is passed through a shaping filter having a transfer function designated H_1 (ω), which delays the waveform one interval and adds it to itself. As a result, the original two-level signal is transformed into a 3-level signal formed by an algebraic addition of the present and previous pulses. For example, if three unipolar pulses are sent in succession at time intervals of T seconds as shown in figure 2, sampling the signal at odd multiples of T/2 seconds (T/2, 3T/2, etc.) allows the generation of pulses using interference from the adjacent pulse.

When pulse No. 1 is sampled at t = T/2, the value of pulse No. 2 at the same instant in time is not zero but is exactly the same as that of pulse No. 1; the same can be said for pulses No. 2 and No. 3 at t = 3T/2. Thus, overlapping of pulses is deliberately introduced at the sampling instants; at the receiving end, the algebraic sum is observed and interpreted. When both pulses are present, as at t = T/2 and t = 3T/2, the value of a sampling instant is $4/\pi$, or twice that of a single pulse. At t = 5T/2, only one pulse is present, so the value is $2/\pi$. Finally, when no pulse is present at t = 7T/2, the value of the sampling instant is zero. Consequently, the signal has three distinguishable levels at the sampling instants.

Limitations on signal bandwidth are not the only bonds on data transmission. As a rule, data communication utilizes existing common carrier telephone plants which were designed primarily for voice communication. The channels provided by these facilities, while adequate for their intended purposes, are afflicted with noise, linear distortion, non-linearities (such as harmonic distortion and spurious signals caused by system compander action), and the abrupt changes in channel characteristics known as phase and amplitude hits. These impairments modify the transmitted data signals, resulting in errors when the receiver reconstructs the message.

Noise in a communication channel is of two basic types: white and impulse — the latter appears as sharp bursts of energy arising from such

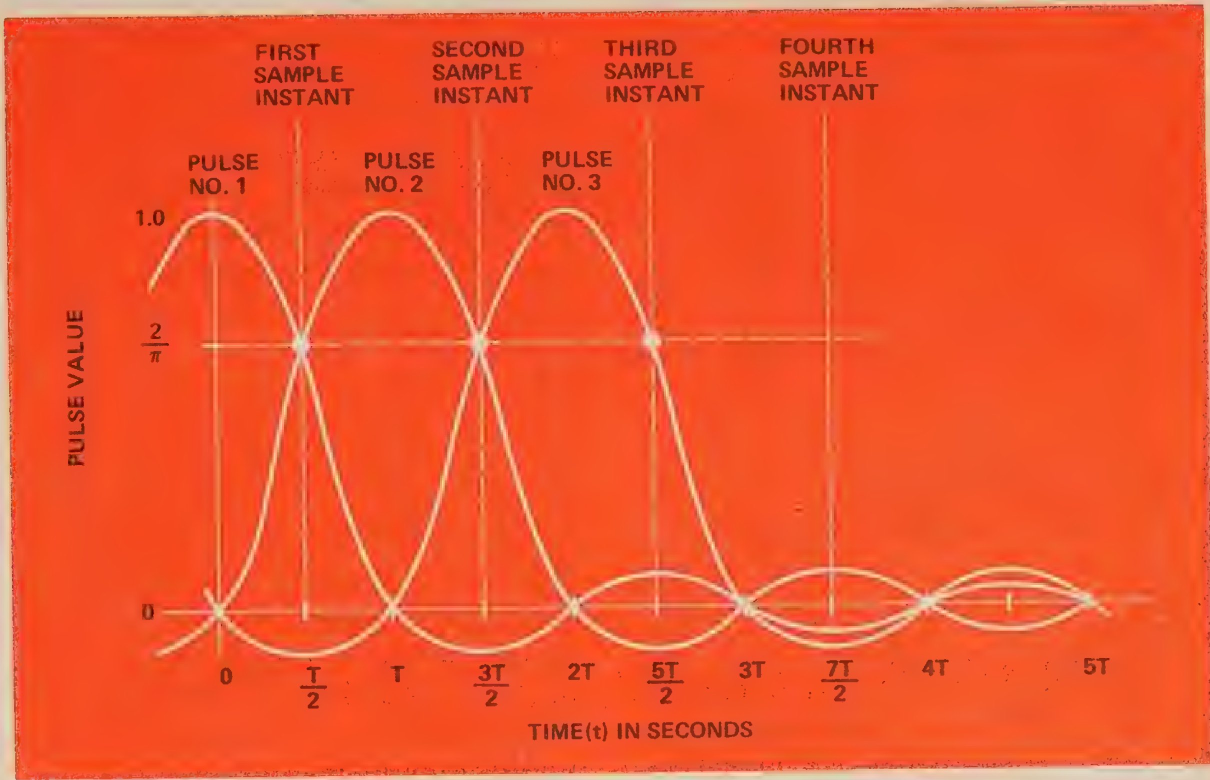


Figure 2. Superposition of pulses in the duobinary coding process results in the creation of a three-level signal with values of $4/\pi$ at sample instants $t = T/2$ and $3T/2$, $2/\pi$ at $t = 5T/2$, and zero at $t = 7T/2$.

sources as electrical storms and switching in telephone plants. White noise interferes more with speech than it does with binary data, because the amplitude of the data pulses can be set above the noise level; however, as the number of signal levels is increased for a fixed power, the noise margin decreases and white noise can become a channel capacity limiting factor. Impulse noise is a much greater problem in data transmission than white noise, because the short duration and high instantaneous amplitude of an impulse spike may be easily interpreted by the receiver as a data signal, resulting in an error in message reproduction.

Equalization

Of most concern in high-speed data transmission is a form of distortion called intersymbol interference, which is defined as the distortion caused by the tails of preceding and succeeding pulses intruding into the time slot of the pulse currently being transmitted.

Intersymbol interference results from the non-uniform delay and amplitude characteristics of communication channels. An equalizer – basically an adjustable filter – is used to introduce a controlled amount of delay at certain frequencies to achieve uniform delay and amplitude over the entire bandwidth. At lower transmission speeds (up to about 2400 bits per second), it is possible to take a statistical sampling of signals sent over the network and determine the average performance of the channel; it is then possible to design compromise equalizers that will compensate for average distortion characteristics. At higher speeds, however, more efficient utilization of bandwidth is required, as is equalization that is more precisely tailored to the particular channel in use during any one random connection. This is beyond the capability of fixed equalizers. To meet these requirements, equalizers have been developed that change their characteris-

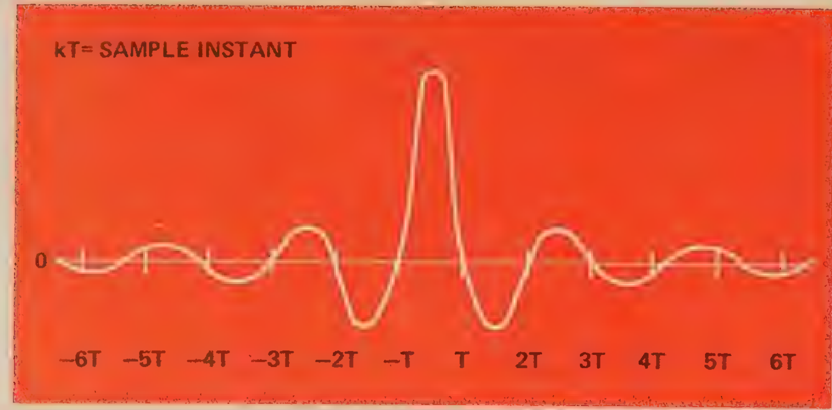


Figure 3. Output of a 6-tap zero-forcing equalizer. Zero-forcing equalization tries to make the pulse cross zero at sample instants corresponding to tap placement (±T, ±2T, ±3T).

tics automatically to fit those of the communications channel.

Automatic equalizers make use of a transversal filter — a tapped delay line whose taps are connected to gain control circuits. By varying the gains of the taps, the controlled amplitude and delay distortion introduced into the signal components can be set to compensate for the effects of the channel. In general, two types of automatic equalization can be distinguished: preset and adaptive. In preset equalization, test signals are used to adjust the tap gains prior to, or during breaks in, data transmission; a major shortcoming of this technique is that it requires the test pulses to be transmitted during a separate training time whenever the channel characteristics change, thus interrupting data flow. Adaptive equalization provides for constant monitoring of the data signal and adjustment of the tap gain settings to give optimum equalization at all times.

In practical adaptive equalization, data transmission is preceded by a short training period, during which test pulses allow determination of an initial equalization level and the setting of the tap gains accordingly. As data are transmitted, samples are taken and compared to the equalization level; any difference between channel characteristic — as revealed by the data sample — and equalized sample results in the generation of an error sample and the setting of the tap gains to cancel the difference.

Phase modulation has proved to be an effective and reliable technique for high-speed data transmission, and adaptive equalization techniques suitable for use with phase-modulated systems are becoming increasingly important in commercial applications. Implementation of these processes depends upon logic circuitry in the equalizer solving simultaneous equations. One procedure, based on a zero-forcing algorithm (an algorithm is a procedure for solving a mathematical problem in a finite number of steps), involves looking at the equalizer output pulse shape; by changing the tap

Figure 4. Output of a 6-tap mean-square equalizer. The pulse is not constrained to pass through zero; instead, the equalizer minimizes the sum of the square of the errors, resulting in less distortion over the entire bandwidth.

gain settings, the equalizer attempts to get the pulse to cross zero at the sample times (see figure 3). If, for example, three taps are placed on each side of the delay line center tap, the equalizer will try to force the output pulses to cross zero at three sampling instants on either side of the main pulse. Under heavy distortion conditions, however, there is the likelihood that, while forcing a certain number of pulse samples to cross zero, the samples not being looked at will be distorted.

Another procedure for equalizing phase-modulated systems was recently developed by GTE Lenkurt, utilizing an incoherent, mean-square algorithm. This technique, which is called incoherent because it does not require carrier phase or amplitude reference information, looks at all of the output pulse samples and attempts to minimize the mean square (rms) error, rather than forcing certain samples to cross zero (see figure 4). Each received signal sample is used to generate a reference to which the next succeeding sample is compared. Due to the use of the previous received samples for both phase and amplitude reference, operation of the equalizer is almost totally independent of frequency offset and phase jitter, and avoids the problem of establishing a reasonable initial phase reference in fast start-up systems. In actual use, this algorithm has allowed the analog components to be replaced by A/D converters and a digital processor.

The common carrier communications systems constitute a vast network of transmission facilities which, while designed primarily to handle speech signals, can be made to transmit digital signals. In transmitting data over such communication systems, accuracy is of utmost importance; the redundancy which allows significant interference in speech is missing in data, and channel characteristics that are acceptable for analog signals may alter a digital signal and produce errors in message reproduction. Solving the problems presented by channel distortion is therefore essential if the most efficient use of transmission facilities is to be realized.

BIBLIOGRAPHY

Cherry, C. *On Human Communication.* Cambridge, Mass.: The M.I.T. Press, 1966.

"Data Communications," *The Lenkurt Demodulator*, September and October 1966.

"Information Theory and Coding," *The Lenkurt Demodulator*, April and May 1965.

Lender, A. *Correlative Signal Processing.* Stanford, Ca.: Stanford University Center for Systems Research, 1972.

Lucky, R.W. "Techniques for Adaptive Equalization of Digital Communication Systems," *The Bell System Technical Journal*, Vol. 45 (February 1966), 255-286.

Lucky, R.W., et al. *Principles of Data Communication.* New York: McGraw-Hill Book Company, 1968.

Shannon, C.E. and Weaver, W. *The Mathematical Theory of Communication.* Urbana, Ill.: University of Illinois Press, 1949.

Tracey, R.J. "An Incoherent Adaptive Mean Square Equalizer," *International Conference on Communications*, (June 1974), 9F-1 – 9F-5.

SECTION V
GENERAL COMMUNICATIONS

GTE LENKURT

DEMODULATOR

MARCH 1972

TWENTIETH ANNIVERSARY

Yesterday and Today

The Demodulator has kept pace with innovations in electronics for 20 years, serving its readership as an informative, educational vehicle, covering some of the most complex advances in the telecommunications industry. The need for this publication arose early in Lenkurt's history as transmission technology became increasingly complex.

Man's progress in the communications field has been significant since Alexander Graham Bell's invention in 1876. The first practical telephone circuits employed a single wire with a ground return and a telephone connected at each end. Under these conditions, each telephone could be connected only with the telephone at the opposite end of the circuit – and not to any others. From this early arrangement there developed the idea of establishing a practical means of interconnecting all the telephones in a local area through an exchange office. Because of excessive electrical disturbances, the single-wire line was replaced with the two-wire line which utilized two closely paralleled wires. One of the wires provided the current return path instead of returning the current through the earth. This arrangement was known as an open-wire line since bare wire was attached to insulators on crossarms that were mounted near the tops of poles.

As the demand for more telephone circuits increased, and the number of wires stretching between cities began to reach its limit (see Figure 1), a method of increasing the number of telephone circuits without adding more wire became imperative. Research in this area led to the development of multiplexing – the combination of several individual voice circuits for simultaneous transmission over a common transmission line. This development existed in a primitive state shortly before 1900. The advent of the vacuum tube and its further refinement by addition of the control grid cleared the way for the rapid advancement of multiplex telephony.

Early Systems

In 1918, the Bell System placed the first carrier system into operation between Baltimore, Maryland and Pittsburgh, Pennsylvania. This system, designated as Type A, provided four two-way channels for use over a single open-wire line, with the same carrier frequencies used for each direction of transmission. The Type B system was put into use by Bell in 1920. In this system, three two-way channels were provided using different frequencies for each direction of transmission. The technique of using different frequencies for each direction provided what is known as an equivalent *4-wire system*. Both the Type A and B systems used amplitude modulation to superimpose the voice signals onto the carrier frequency. In the Type A system, the carrier and one sideband were suppressed with only the remaining sideband being transmitted. In the Type B system, one sideband was suppressed while the other sideband and the carrier were transmitted.

The Type C system, introduced in 1925, incorporated the best features of the two earlier systems. This system

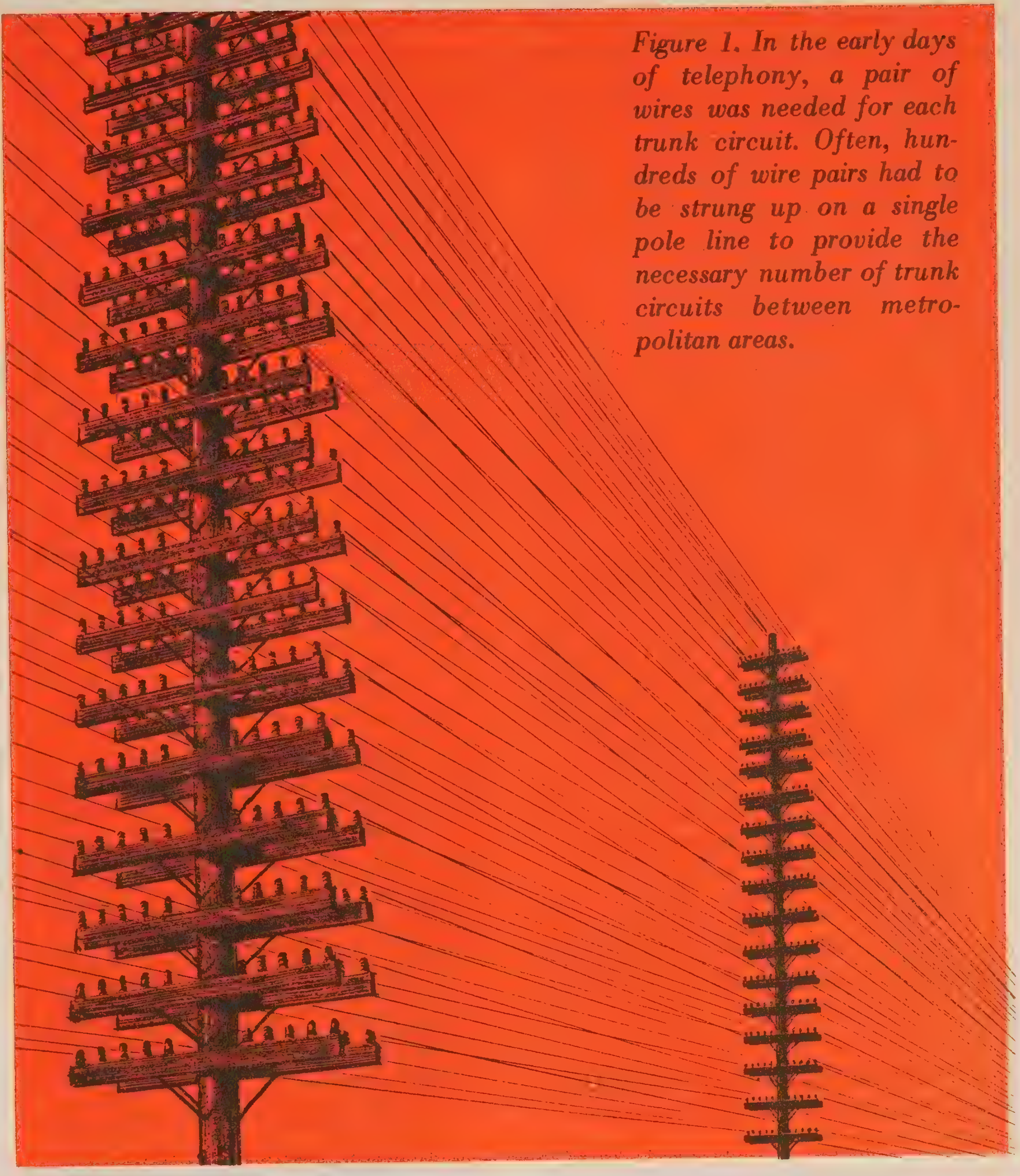


Figure 1. In the early days of telephony, a pair of wires was needed for each trunk circuit. Often, hundreds of wire pairs had to be strung up on a single pole line to provide the necessary number of trunk circuits between metropolitan areas.

provided three channels using different frequencies for each direction of transmission, and transmitted only one sideband. The Type C system became extensively used throughout the Bell System's long distance toll routes and by 1928, several transcontinental 3-channel routes were in operation in addition to many shorter systems between such points as Chicago and Pittsburgh, and San Francisco and Los Angeles.

12-Channel Systems

Technological advances soon proceeded far enough to permit the development of 12-channel carrier systems designed to operate over cables and open-wire lines. A 12-channel system was used on a transcontinental cable route in 1938. The frequency band for this system was from 12 to 60 kHz for transmission in both directions. This was done by using a different wire pair for each direction, thus establishing a

physical 4-wire system. The lower line frequency was achieved by using a new technique called *group modulation*. In the earlier systems, the carrier line frequencies were accomplished by a single direct modulation step. Group modulation, however, used two or more steps of modulation to establish the line frequencies. One of the most significant advantages of group modulation was that it provided a simplified means of interconnecting standard sub-groups of channels at line frequencies, a technique that was to be employed extensively in later carrier systems.

Advances in Communications

As World War II became a reality for the United States, the telecommunications industry, along with most other industries, curtailed its peacetime endeavors and concentrated its resources on the war effort. During World War II great advances were made in the field of military communications. These advances were later to have a significant influence upon commercial communications such as microwave, video and higher-density carrier systems.

In 1944, with the end of the war somewhat in sight and with the relaxing of commercial development restrictions, men with foresight realized that there would soon be a demand for expanded communications facilities. Two such men were Len Erickson and Kurt Appert — founders of the company whose name still bears the combination of their first names. Founded in 1944, Lenkurt Electric was destined to make significant contributions to the telecommunications field. With their experience in military communications, Erickson and Appert began to design carrier equipment that would be suitable for commercial purposes, and in 1945 introduced the Type 12 and 17 carrier systems. Both systems were single-channel systems and collectively utilized a frequency range from 12 to 20 kHz, with signaling above the voice channel. These systems were designed for use on open-wire lines, and repeaters were available to permit use on short, medium, or long-haul circuits. The typcial appearance of these early units is shown in Figure 2.

The Demodulator

By 1952, the field of telecommunications had achieved a high degree of sophistication. Innovations such as transistors, silicone diødes, and printed circuits began to be seen as practical applications. As equipment became more specialized, so did the men who

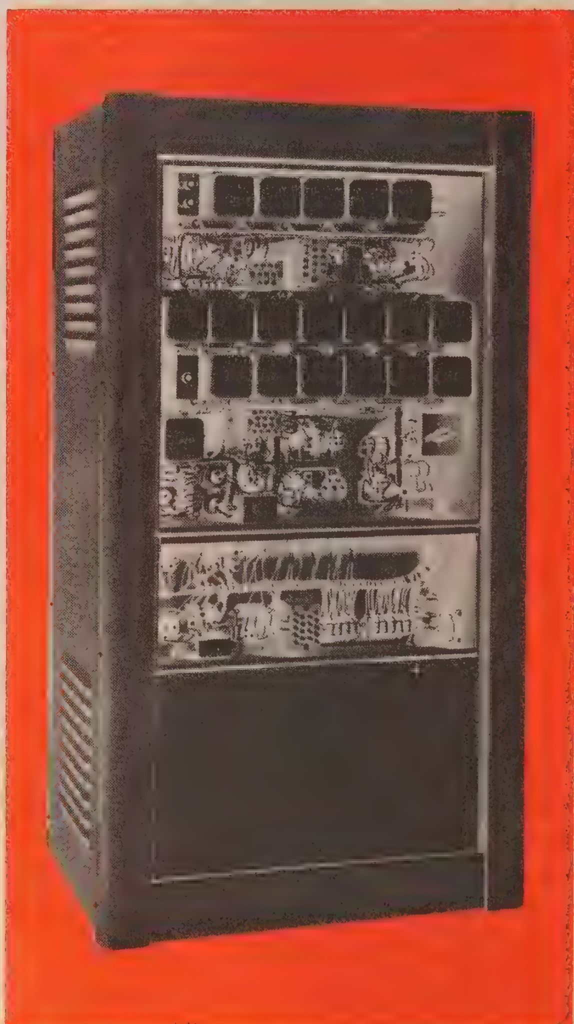


Figure 2. Typical appearance of early Lenkurt Type 12and Type 17 terminals.

designed it. In many cases specialists in one field were only vaguely aware of what was taking place in other fields of communications. Clearly, a method was necessary for acquainting engineers, managers, technicians, installers, and other individuals who worked in the communications industry, with the various advances being made. It was in this light that the Demodulator came into existence, and its first issue was published in March, 1952.

The early years found the Demodulator staff busy with explanation of basic telecommunications concepts and their applications. Even as these concepts were coming to light, new ones were being introduced, and throughout its 20-year existence, the Demodulator was to print information on such topics as open-wire carrier, signaling, cable carrier, microwave radio, data transmission, and many other related products and applications.

In a sense, the Demodulator has been both witness and recorder to many of the developments in the telecommunications field. In the past 20 years, the Demodulator has endeavored to report not only on GTE Lenkurt developments but also on advances which have taken place in the telecommunications industry as a whole.

Plug-In Components

A significant contribution was made to the telecommunications industry in the early 1950's with the introduction of the "universal carrier" concept. In this Lenkurt development program, carrier systems for all the different transmission mediums were developed by using the same basic parameters for all systems. This eliminated the necessity of demodulating the line signals to voice level at the interconnect point (between cable, open-wire, or radio systems), with only the line and common equipment required at this interface. The result was a substantial cost savings to the user and significantly better transmission quality, since two steps of modulation were saved at each interface. Variations between systems were made only when required in order to obtain the necessary characteristics peculiar to a particular system.

The basic chassis for the universal carriers was the same throughout all systems regardless of channel assignment or particular system, while the frequency-determining components, such as filters and oscillators, plugged into specified pregroup arrangements. An example of the plug-in concept is shown in Figure 3.

Miniaturization

One of the developments which the Demodulator has watched with keen interest has been in the field of miniaturization.

The electronics industry has undergone a progressive miniaturization of equipment component parts as well as entire electronic circuits. From the early vacuum tube hand-wired circuits, the progression has been to transistorized printed circuits, to thick and thin-film integrated circuits.

As reported in the November and December 1971 issues of the Demodulator, the miniaturization of filters has been dramatic. From the early large and bulky filters, technology in this field has improved greatly with the introduction of the polylithic filter developed by GTE Lenkurt. The relative size-reduction of filters over the years is shown in Figure 4. Crystal filters have not only increased the efficiency of equipment, but have also allowed engineers to decrease the size of telecommunications equipment and at the same time increase its performance and capacity.

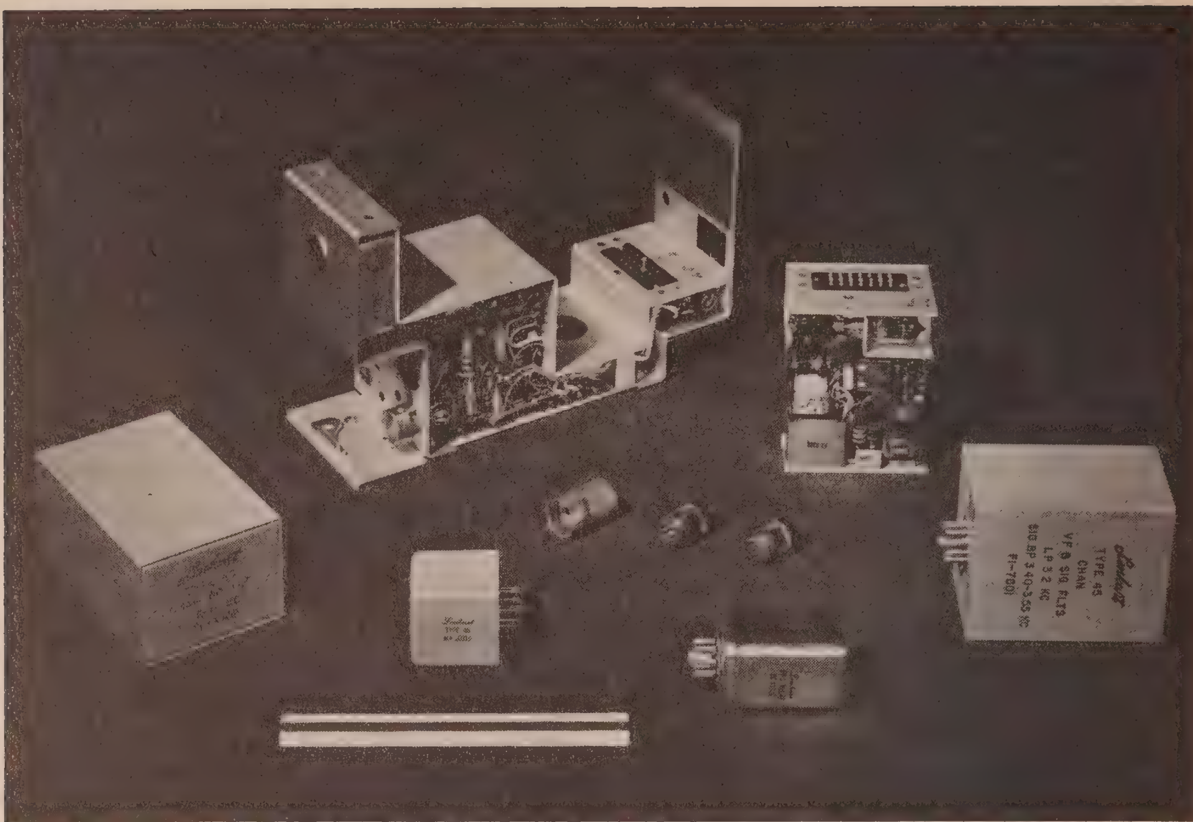


Figure 3. Plug-in components were essential to the universal carrier concept as shown by this Lenkurt Class 45 channel unit.

The major impetus to miniaturization of electronics equipment has been the development of devices and circuits which operate using principles of applied solid-state physics. Because of the importance of solid-state development to the telecommunications field, this subject has been the topic of several Demodulator articles. These articles have progressed in their description from basic transistor analysis to principles of integrated circuitry.

Multiplex Technology

The advances in multiplex technology are as impressive as those made in other fields of communications. In 1946, Lenkurt was producing 3- and 4-channel systems suitable for open-wire lines. These systems required that all channels, common equipment and repeaters be installed in order to function properly. A major development of the late 1940's was the introduction of fully "frequency-stackable" equipment. This type of equipment allowed all channels to operate independently with the exception of the power supply and line filters. Also, the repeaters for the equipment were stackable in that any channel and its corresponding repeater could operate independently without requiring the installation of repeaters for the other channels.

From these early beginnings, the Demodulator has observed and reported the great parade of technological advances in the multiplex field. Developments in solid-state circuitry and filter design have dramatically reduced the size of equipment and expanded channel capacity to such a degree that the space required for a 1260-channel system is not much more than that required for an early 48-channel system (see Figure 5).

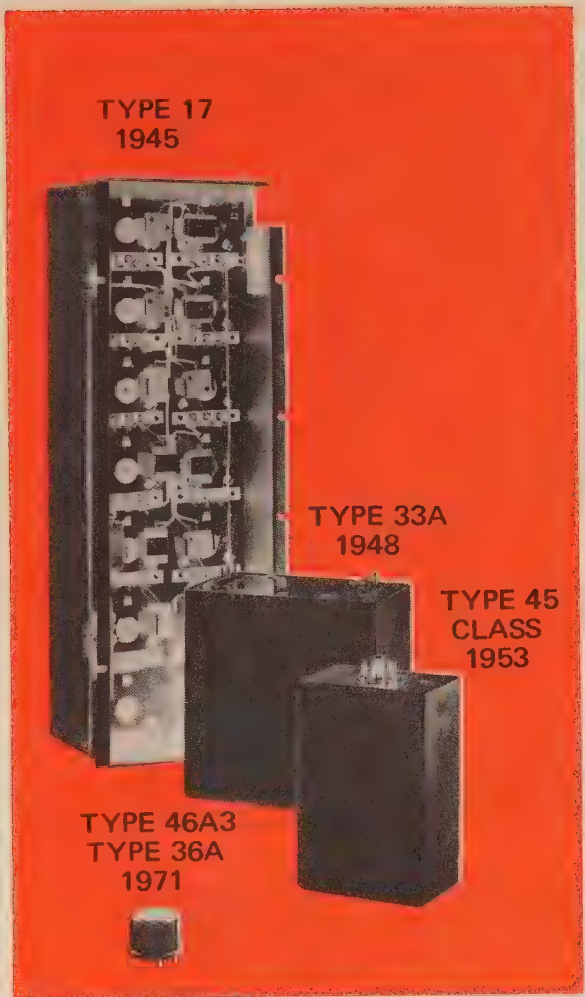


Figure 4. Filter size in GTE Lenkurt equipment has changed dramatically over the years.

Data and PCM

With the advent of the computer age, the Demodulator witnessed and recorded some of the great advances in digital technology. One of these advances was the duobinary coding technique of data transmission which was developed by GTE Lenkurt, and which proved to be a breakthrough in the world of digital transmission. Using this coding technique, which in effect doubled the capacity of the binary system, data transmission systems such as the Lenkurt 26C, could transmit digital signals over conventional voice channels by wire, land or submarine cable, or microwave radio, at fixed speeds of 2400 bits of information per second. The duobinary method, explained in the February, 1963 issue of the Demodulator, allows higher data speeds to be achieved with simple frequency modulation techniques rather than the complex methods previously required.

PCM (pulse code modulation) was first discussed in the Demodulator in January, 1959. The telecommunications industry was long since aware of the advantageous properties of PCM – relative freedom from noise and interference, and regeneration of signals without significant distortion. However, it was not until solid-state technology had made significant advances that PCM became a practical means of transmission.

Unique in its concept, PCM differed greatly from conventional FDM (frequency-division multiplex) techniques. Like FDM, PCM channels share the same transmission medium. Unlike FDM, all PCM channels are not transmitted simultaneously, but "take turns," each being connected to the line very briefly, then replaced by the next. Since PCM is a relatively new form of information transmission, and since the future of PCM as an established transmission technique is apparent, the Demodulator has devoted several articles in recent years to the theory of PCM and its applications.

Microwave

As early as May, 1952, the Demodulator became involved in the explanation of microwave transmission. Using the Lenkurt 72A microwave system as an example, the Demodulator confronted the problems of transmission losses in radio links. In these days, the microwave frequency range was around 900 MHz with a transmission capacity of only up to 36 toll-quality, carrier-derived voice channels, with repeater spacing of 25 to 35 miles.

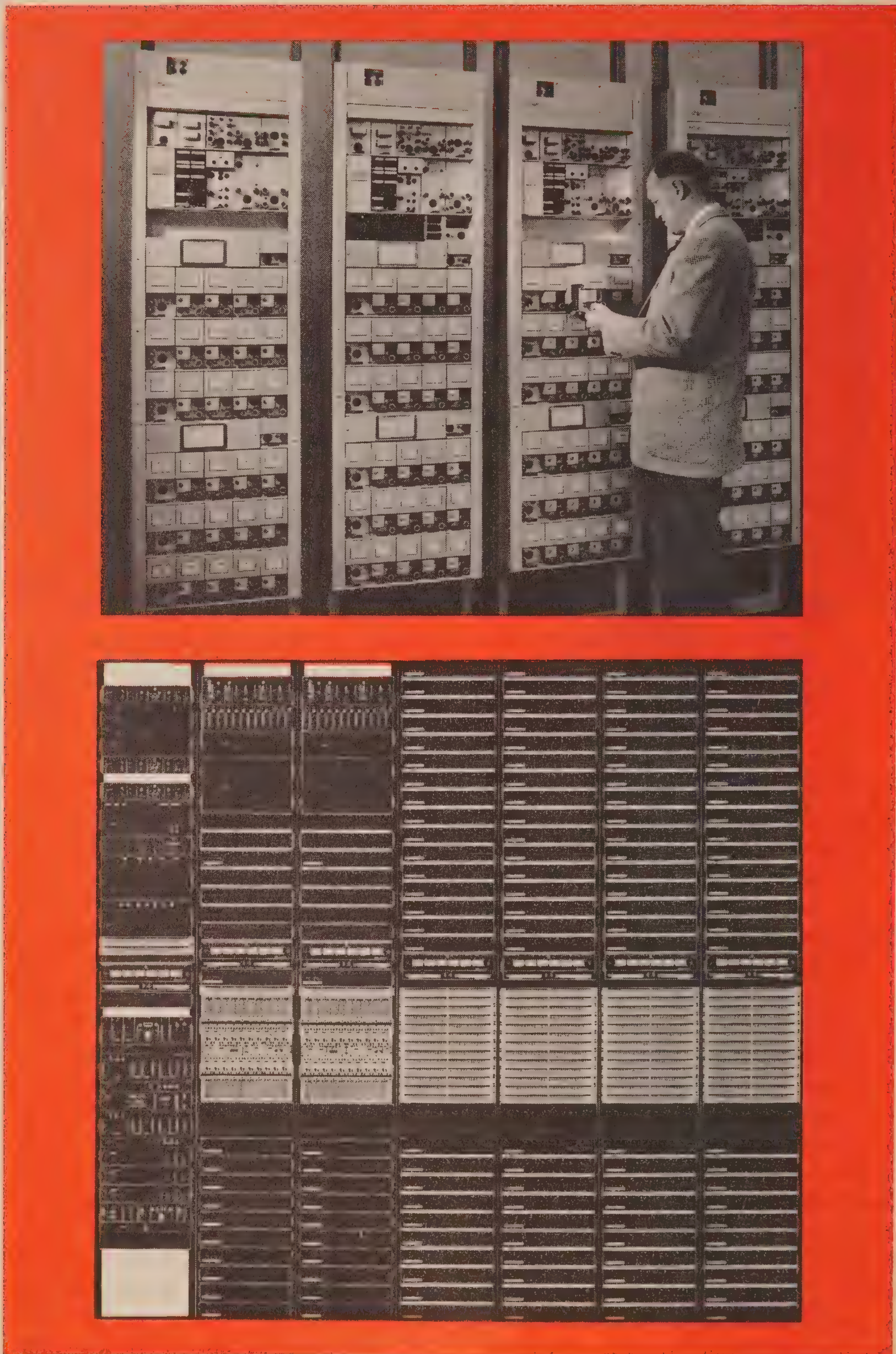


Figure 5. The 45 BX multiplex terminals were large compared to modern 1260-channel systems such as Lenkurt's 46A3.

Methods of establishing microwave paths have varied over the years. One method, employed by the colorful "Captain Eddy," utilized an airplane equipped with low-altitude radar altimeters. Captain Eddy's low-flying aircraft could often be seen skimming the ground in order to establish ground clearance for microwave transmission paths.

Although line-of-sight radar mapping is still in use today, the most widely used method is path layout using topographic maps, confirmed by ground surveying. Verification of the ground survey includes the use of strobe lights, altimeters, mirrors, and/or metallic reflectors. Another method utilizes a helium-filled balloon with a light source or reflector attached to it. A *simplified* example of this procedure goes something like this: During calm weather, an individual at point A allows the balloon to ascend, controlling the ascent by means of a line attached to the balloon. An individual at point B watches with binoculars in the direction of point A. When point B observes the light source or reflector, he notifies point A. Point A then takes an altitude reading which will indicate the obstruction height of this particular microwave link.

As microwave technology advanced, telephone companies began to find that microwave systems had a definite place in their outside plants since numerous installations had demonstrated that microwave systems could be engineered to be equally, or more reliable than conventional open-wire lines or cables. Subsequent Demodulator articles considered such topics as microwave propagation techniques, waveguide, and antenna applications.

Continual improvements in design and performance — especially solid-state advances — greatly increased the popularity of microwave systems so that by the early 1960's, there existed some high-quality systems capable of handling more than 1200 voice channels or one video channel. Today 1200-1800 channel systems are a matter of course, and their 36-channel predecessors seem to be almost forgotten in the mass of telecommunications history. However, these first systems, their problems and applications, are periodically recorded in the 20-year span of Demodulator issues.

Coaxial Cable Systems

Several Demodulator articles over the years have offered coaxial cable systems as a logical alternative in areas where microwave frequency bands are becoming increasingly congested.

Although coaxial cable has been used for communications since 1936, installation costs have limited its application. Some advantages of a coaxial cable system are high reliability, low maintenance costs, and low-cost expansion. The increasing demand for instant communication between individuals or computers, and the corresponding bandwidths necessary (high-speed data, video, video telephone), often justify the installation of coaxial cable. One coaxial cable link may handle as much traffic as several microwave links since each microwave channel has definite bandwidth limitations.

With the rapid growth of big cities, which is where most high-density communications links terminate, coaxial cable systems are becoming increasingly attractive as so-called "entrance links." This implies that a microwave system does not extend to the downtown area, but the baseband signal is brought there from a point on the perimeter of the city by coaxial cable. This prevents having to re-engineer a system every time a new high-rise

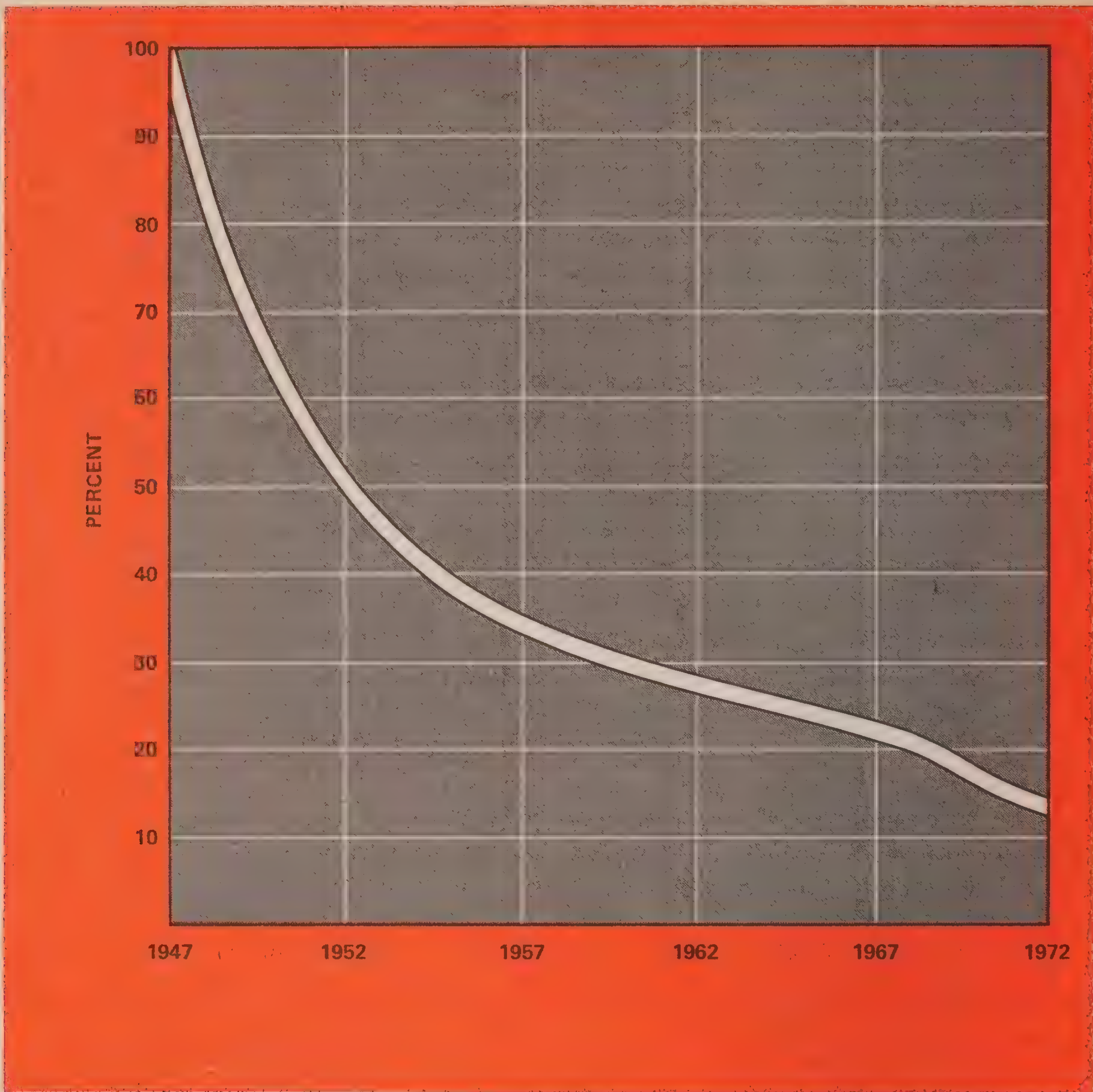


Figure 6. The decreasing price trend of GTE Lenkurt carrier and multiplex equipment reflects the general trend of the telecommunications industry.

building is erected in a microwave path. Also, this method avoids the very real problem of not being able to obtain a microwave frequency in the desired band.

Exchange Carrier

The October, 1960 issue of the Demodulator presented the difficult problem of exchange carrier and offered a solution. Long distance systems required great care in regulating and reducing distortion. For this reason, carrier equipment designed for toll or long haul service was quite elaborate and costly. For shorter distances, the same complex terminal equipment proved too expensive, especially since shorter systems didn't require the same features found in a long system.

Using the Lenkurt 81A exchange trunk carrier as an example, the Demodulator described the advantages of this type of system. The main overall advantage of exchange carrier is its flexibility in expanding plant facilities to meet new demands. As cable cir-

cuits become fully utilized, exchange carrier channels can be added one at a time. This provides a means for orderly expansion even in areas where demand becomes much greater than had been expected.

Prior to the introduction of exchange carrier, the cost of carrier equipment was such that it could only prove-in on long distance routes. Exchange carrier provided exchange and toll-connecting trunk carrier service for routes of up to 30 miles.

Subscriber Carrier

A particularly successful endeavor for GTE Lenkurt was the development of the subscriber carrier in 1967. Carrier transmission over open wire or cable between the central office and a subscriber is the function of subscriber or station carrier. An ordinary cable pair carries one voice or data channel. This same pair, using a subscriber carrier system, can carry from one to six frequency-derived circuits which can be used for voice or data.

As pointed out in the March, 1970 issue of the Demodulator, subscriber carrier transmission can be an economical alternative to cable when the annual circuit requirements are small or where the circuit length is quite long. If future requirements are uncertain, subscriber carrier provides interim relief, allowing time to gather reliable data before making a major cable addition. Subscriber carrier systems are also advantageous where new cable additions would provide a great many extra pairs that would be idle for several years.

Cost Reduction

In the last two decades, manufacturers of telecommunications equipment have marketed many different types of voice carrier systems ranging from relatively simple single-channel products to highly complex 1260-channel systems.

While the complexity of each carrier product has greatly increased, the average price per channel has decreased steadily over the years and current products are priced substantially less than their representative predecessors of twenty years ago. The main factors which have been responsible for price reductions during the past two decades have been the technological developments in the state of the electronics art and improved design and manufacturing techniques. Using 1947 as an index, figure 6 shows the average declining price trends of GTE Lenkurt carrier and multiplex equipment.

The Future

In the telecommunications industry, a backward glance at the past is permitted only occasionally (*such as on anniversaries or birthdates*). The usual point of attention is to the unlimited possibilities and applications that lie in the future. Challenges are being met daily to keep products ahead of technological changes, and at the same time new demands are being met with more sophisticated communications systems.

As new developments take place, the Demodulator will record and report on these events, for the success of the Demodulator can only be measured by its ability to bring to its readers available information on current developments in a concise and readable form.

The Lenkurt DEMODULATOR

JANUARY 1971

communications and the environment

Proper application of today's communication technology can bring people together and improve the atmosphere in which they live.

Environment includes not only geographic features, but also the people and the subsequent culture of an area. Present communication links can be expanded for environmental channels – voice and video channels for education and exchange of ideas, as well as data channels for earth resources management. This expansion can be realized by utilizing today's communications technology.

Remote data collection and centralized computer analysis of the data can provide an efficient means of measuring, analyzing, and correcting environmental pollution. By providing more channels of communication, more opportunities for expression of ideas through dialogue would be available. These communication channels can be provided by increased two-way video, voice, and data communication.

Pollution Control

Although it is not physically or financially feasible to establish manned laboratories in every geographic location where pollution is most likely to occur, it is possible, by means of a network of unmanned data collection stations, to sample the surroundings and transmit information on air, earth, and water conditions to a central processing laboratory for analysis. In this way, computer technology and remote data acquisition can contribute to pollution control.

Prototype pollution monitoring systems are presently in operation. What look like ordinary navigation buoys are really ocean pollutant detectors. Instrumented buoys, anchored in oceans and inland waterways are equipped with sensor systems and automatic data handling equipment. These unattended buoys are able to measure and transmit such data as water and air temperature, wind speed and direction, and barometric pressure. Such systems are being designed for low-power consumption and long-life expectancy which should provide easily-maintained, low-cost environmental monitoring. A network of ocean monitoring buoys, or stations can communicate with a central processor either over a direct microwave link or via satellite relay links (see Figure 1).

Another pollution detection device now under development employs a patrol aircraft that measures the changes in microwave radiation from the surface of the water (see Figure 2); thereby, determining what the pollutant is – oil or gas – and how thick the spill is.

Similar tests can also be made on the atmosphere to detect air pollution. The proper transmission links permit measurement at many remote locations and processing at one location. Depending upon the results of the analyzed data, the proper corrective

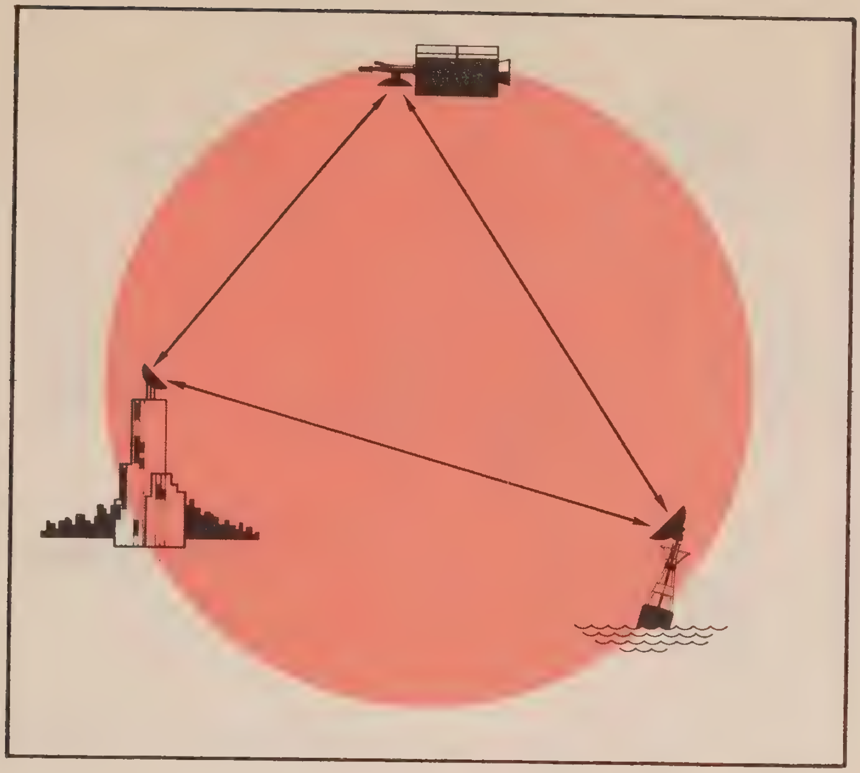


Figure 1. Satellite relay techniques are used to monitor natural resources when a direct microwave link is not practical or feasible.

actions can be transmitted by the central processor for the particular pollution location.

Information can be transmitted from the data collection points to a central processor by microwave techniques. For getting information from the remote data collection points, satellites seem to offer a convenient means. In some cases, depending upon the type of data being collected, the satellite may be able to actually gather the raw data and transmit it directly to the central processor without a surface-collection system.

Satellite Network

A network of satellites and surface-probing sensor systems may be used to study natural resources. In addition to the oceans and air, this network can take inventory of what, where, and how well forests and crops are growing, and the condition of the soil and its ability to be put to work; thus permitting regional, national, or global predictions of crop yields, livestock inventory, and patterns of fire, insect, and disease damage. Information about stream and river flow, excess surface water, pollution, and glacial action can be studied in order to plan better irrigation and flood control systems, develop and maintain water resources, and control erosion.

Air pollution is generally correlated with population distribution and geographic features that can be studied with satellite mapping techniques. Detailed maps of the earth's features can be used for planning land use, urban development, and transportation facilities. Aerial data collection can also be used to map ocean currents, ice, and other navigational hazards. Fish and other marine biology of interest, as well as pollutants, can be studied for the seafood industry, shipping, and marine ecology.

Surface-collection relay satellites and remote-sensing satellites, along with non-satellite remote sensing devices — including sounding rockets, balloons, aircraft, buoys, and ground-based platforms — are capable of transmitting the gathered information

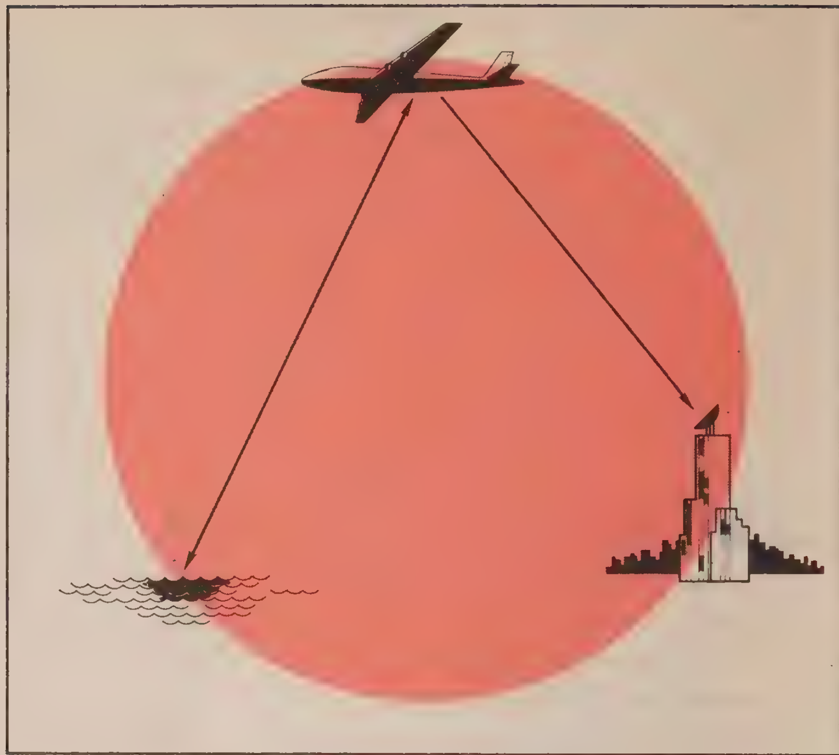


Figure 2. Airborne systems detect petroleum spills in the ocean and transmit the information to a central office for cleanup operations.

to a central computer. The computer's role in this overall environmental management system is that of soothsayer – if, for example, a decision were made to irrigate thousands of square miles of desert to create a new agricultural area, the computer could predict such things as the plan's effect on: climate, population, water resources, and international trade.

In order to manage world resources effectively, adequate information must be available. Information has for centuries been gathered by man on the surface of the earth. In recent times, aerial observations have broadened the field of view, the amount, and the usefulness of the information. With the mass acquisition of data and sophisticated computer processing, it may be possible to stem the tide of diminishing resources, and pollution of the existing resources.

Human Environment

Solving the problems of an area's pollution and diminishing natural resources will do little to improve the total environment, if the people in the area are unable to communicate and clear up differences. These differences often represent a widening gap between expectations, and reality. In an affluent society, we expect more, and better communications are raising these expectations. Through proper education and exchange of ideas it is possible to bring expectations in line with reality.

The areas of communication offered to bring expectations closer to reality include: education, community expression, cultural enrichment, and politics. Some specific services offered include: home library service, facsimile, delivery of mail, crime detection and prevention, remote data acquisition and central processing, educational television, remote participation at conferences, and armchair shopping.

Expanded Services

These new services can be divided into two classes: one-way transmission with no interaction between transmitter and receiver; and two-way trans-

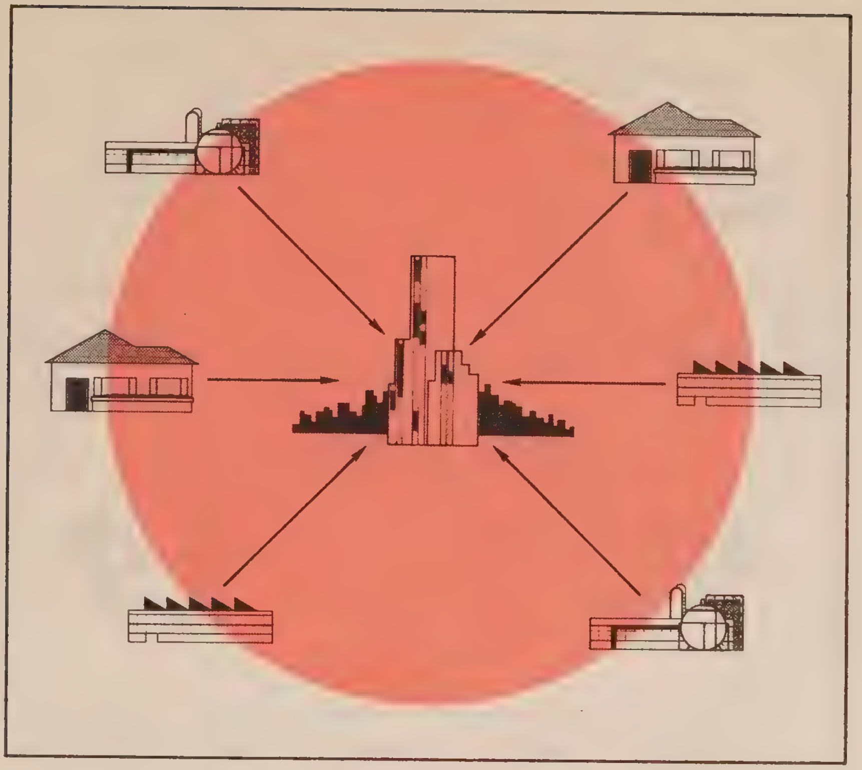


Figure 3. Utility meter-reading employs one-way transmission from many subscribers to one central office.

mission where there is a transmitter and receiver on both ends which provides the opportunity for interaction and response.

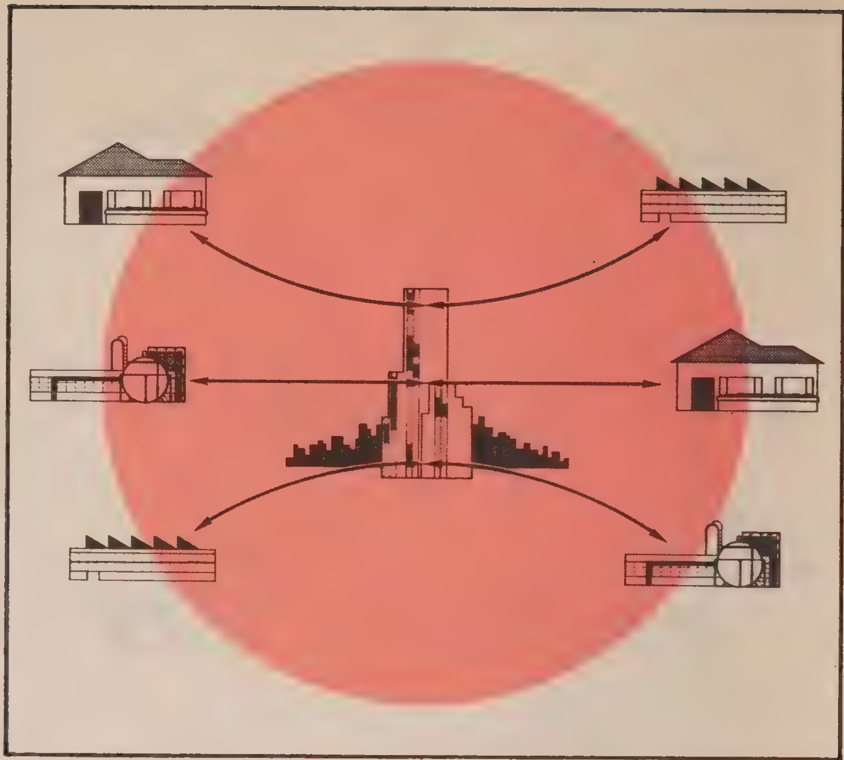
Utility meter-reading is one-way transmission from many subscribers to a central office where the information is processed (see Figure 3). The gathered data from each subscriber is sent through a central processing unit for charge computation. The actual billing could be included in the processing, which would make meter-reading a two-way transmission process. But, it is more likely that billing will continue to use a centralized mail distribution system, since it would not be economical for utilities to operate their own video or data transmission system.

Facsimile (the art of sending pictures or other printed material) is a form of one-way transmission in that there is no interaction between the transmitter and receiver, but both terminals are transmitter/receivers. As technology advances, it may eventually become economical to bring facsimile into the home for such things as home library service and newspaper distribution — if a printed copy of the transmitted image is desired. The transmission of color is possible as demonstrated by color television, but a color facsimile printout device needs to be perfected. Law enforcement agencies are using black and white facsimile printers to speed information across the country for crime detection and prevention. The addition of color would offer improved image recognition.

If printed copy is not needed a video system like television provides readable, although not permanent, written material. The information is read directly off the screen and when finished, the viewer terminates the signal. Cable television, with local programming, could provide channels to bring these services — library and newspaper — into the home. Videophone service could also bring these visual images into the home.

Mail transmission and distribution, as well as video-phone, is a two-way transmission service that could use

Figure 4. Mail transmission and distribution, as well as video-phone, uses a switched network.



the same transmission and distribution plan that is presently used for telephone service. That is, a switched network where an individual sends his message through a central office which redirects the message to the receiver (see Figure 4). With mail transmission and distribution, the service need not be completed at the same time; therefore, delaying the interaction or response. This delay would provide for more efficient use of the transmission channels — transmitting mail in non-peak hours. Mail transmission and distribution will not eliminate the letter carrier, but it can relieve the letter carrier of over 75% of his load without transmitting actual correspondence — personal, business, and government letters — over the air or through a cable. The receiver for such a system could be either a facsimile printer or a video screen depending upon whether printed copy is needed for future use. Another plan gives the sender a choice of transmission modes — instantaneous transmission over telegraph lines to the receiving "post office" where a letter carrier would deliver the message or letter "posting" common today where the original document is "hand" carried to its destination.

Totally automated system monitoring is a two-way system using programmed transmitter/receiver terminals. This is essentially the same technique used for natural resource control, but also used for monitoring other remote systems. Remote data access and central processing also includes time-sharing computer service. As the complexity and cost of these terminals is reduced, more people will take advantage of the benefits offered.

Educational television and remote conference participation are similar to video-phone with instantaneous voice and video communication. Where these services differ from video-phone is that there is a central transmitter/receiver and many remote transmitter/receivers that interact with the central unit (see Figure 5). Using such a service, government officials have direct contact with their constituents. This service has the greatest potential

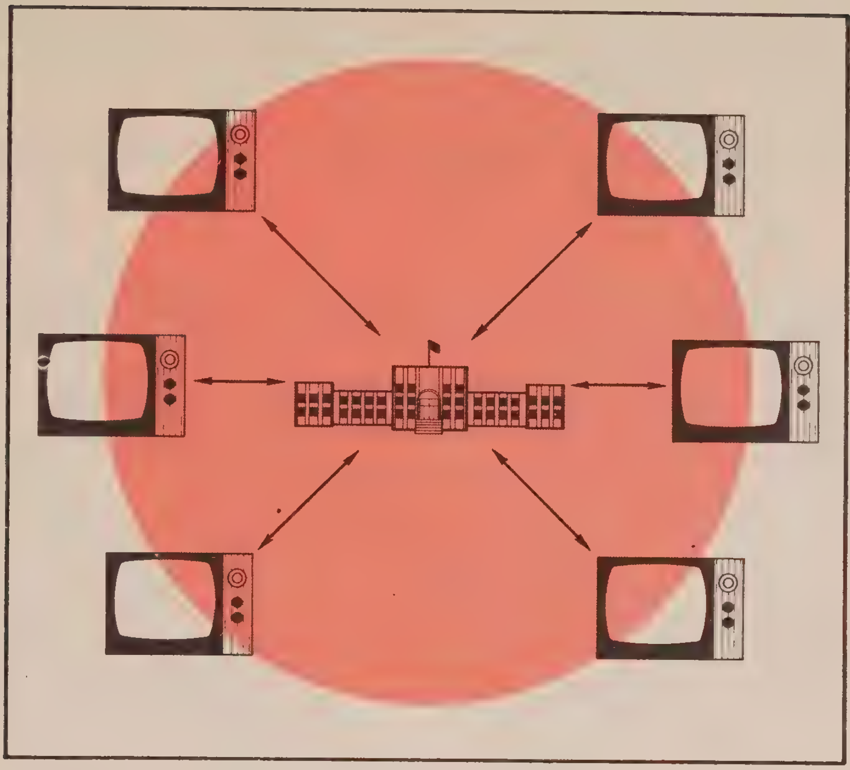


Figure 5. There is two-way communication between one central location and many remote subscribers for educational television and remote conference participation.

for bringing people together because it is possible to clear up any misunderstandings that might arise before they have a chance to cause dissention in the ranks. This service could put expectations and reality into proper perspective. Communities can express themselves over a two-way voice/video channel so the public has the opportunity to know the full story and to express their approval or objection. And, educational television provides the means for educating large masses in one geographic region or select groups scattered over several regions. Increased educational facilities provide the means to close the gap between expectations and reality.

New Direction

Expanded means of communication have the potential to provide a more efficient society with an informed public living in a healthy, plentiful environment. Presently, the possibilities are practically unlimited, but so are the possibilities for this expansion getting out of control. If the best interests of the public are to be realized, the most efficient and most economical systems must be put into effect. None of these expanded services will be totally adopted unless present costs can be substantially reduced. Technology has developed these services, economics will dictate their future.

GTE LENKURT

DEMODULATOR

APRIL 1974

A Glossary Of Signaling Terms

A Glossary Of Signaling Terms

Telephone signaling is the process by which a caller on the transmitting end of a line informs a particular party at the receiving end that a message is to be communicated. But signaling is also a two-way process which includes supervision. Supervision informs the caller that the called party is ready to talk, that his line is busy, or that he has hung up. Supervision is also that part of signaling which holds the voice path together while the conversation goes on.

Even the earliest telephones had to have some way of signaling parties at the other end. As the telephone industry grew, and equipment became more complex, new signaling methods had to be designed to complement the new equipment. This, coupled with the fact that an increasing number of communications equipment manufacturers were designing their own type of signaling equipment, made the subject more confusing. This glossary of signaling terms was prepared to aid the reader in understanding some of the most commonly used words in signaling technology.

ac-dc ringing. A type of telephone ringing which makes use of both ac and dc components—alternating current to operate a ringer and direct current to aid the relay action which stops the ringing when the called telephone is answered.

address. (Sometimes referred to as "called number".) That group of digits which makes up a telephone number. For example, an address may consist of area code, central office, and line number.

AMA (automatic message accounting). An automatic recording system which documents all the necessary billing data of subscriber-dialed long distance calls.

ANI (automatic number identification). Automatic equipment located at a local or toll dial central office used to identify the calling number in customer dialed toll calls. The identity of the calling number is transmitted to CAMA by means of multifrequency pulses that are sent over the same trunk after dial pulsing has taken place.

answer signal. A supervisory signal (usually in tne form of a closed loop) from the called telephone to the central office, and back to the calling telephone (usually in the form of reverse battery) when the called number answers.

audible ringing tone. That tone received by the calling telephone indicating that the called telephone is being rung (formerly called ringback tone).

battery. Usually refers to the dc power source located in the central office. Nominally –24 or –48 volts.

battery and ground signaling. A type of loop signaling designed to double the available signaling current by using battery and ground at both ends of the loop, but with opposite polarities at

each end. This technique doubles the signaling range, but also doubles the impulse noise.

bridged ringing. Any system where all ringers on a line are connected across that line. To avoid shunting of the dc component, a capacitor is placed in series with each ringer.

busy signal. (1) Audible and/or flashing signal, usually 60 impulses per minute (IPM), which indicates that the called number is unavailable, (2) a signal transmitted at 120 IPM which indicates that all voice paths are temporarily unavailable.

CAMA (Centralized Automatic Message Accounting). An automatic message accounting system which is located at a central office, but which serves various adjacent central offices. Calls not processed by ANI (automatic number identification), must be routed through an operator who dials the calling number into the equipment.

carrier signaling. Any of the signaling techniques used in multi-channel carrier transmission. The most commonly used techniques are in-band signaling, out-of-band signaling, and separate channel signaling.

CCIS (Common Channel Interoffice Signaling). A type of signaling system in which all of the signaling information, including supervision and address signals, for a number of interoffice trunks is encoded and transmitted over a separate signaling link by means of time division multiplexing.

CCSS (Common Channel Signaling System). A GTE Lenkurt two-way signaling system which, in telephone usage, carries 2-state signaling, such as E and M or dial pulsing, for 24 voice channels over a separate signaling link. CCSS should not be confused with CCIS, which utilizes data processors, TAC's (Terminal Access Circuits) and other control circuitry not found in CCSS. The CCSS can also be used for low-rate (15 pps and lower) data transmission such as burglar alarms and supervisory and control systems.

class of service. The categorization of telephone subscribers according to specific type of telephone usage. Telephone service distinctions include, for example, rate differences between individual and party lines, flat rate and message rate, and restricted and extended area service.

code ringing. The selective alerting of telephone subscribers on multiparty lines by combinations of short and long rings.

coin-collect tone. A low tone which informs the originating toll operator that the change for a call has been collected by the local operator or the coin control circuit.

coin-denomination tones. The tones produced by two gongs in multislot coin telephones, when nickels, dimes, and quarters are deposited. The tones are detected and transmitted to the operator so that the correct amount can be checked.

coin-return tone. A high tone which informs the originating toll operator that the change for a call has been returned by the local operator or coin control circuit when the connection is not completed.

common battery. A dc power source in the central office that supplies power to all subscriber stations and central office switching equipment.

common-battery signaling. The method by which supervisory and telephone address information is sent to a central office by opening and closing the circuit at the telephone, i.e., depressing and releasing the switch on the cradle of the handset.

common control office. Any of three types of central office—electronic, common control step-by-step, or crossbar. Any common control office receives dial pulses or dual tone multifrequency (DTMF) signals from calling subscribers, or dial pulses, revertive pulses, or multifrequency signals from other offices.

composite signaling (CX). A dc signaling system which requires a single line conductor for each signaling channel, and which provides full duplex operation. In this system, voice frequencies above 100 Hz are separated from the signaling currents by a filter network known as a composite set. Two composite signaling channels are derived from one pair of wires, and four from a phantom group. Composite signaling channels may also be used for dc telegraph or teletypewriter circuits.

cord circuit. The term applied to a circuit that operates by means of a manual switchboard to make connections to subscriber lines and trunks. A three-conductor cord is used with designations of tip, ring, and sleeve used for the conductors.

crossbar switching system. A method of switching which when directed by a common control unit, will select and close a path through a matrix arrangement of switches.

decimonic ringing. A type of party line selective ringing which uses ringing frequencies of 20 Hz, 30 Hz, 40 Hz, 50 Hz, and 60 Hz.

delay-pulsing signal (delay dial, stop dial). An off-hook signal from the called end of a trunk, which is sent to the calling end of a trunk, to indicate that it is not ready to be pulsed.

DDD. Abbreviation for direct distance dialing. Subscriber dialing over the nationwide intertoll telephone network.

dial-normal transmission signal. A secondary dial tone which is returned to an operator to indicate that the rest of a number may be dialed.

dial off-normal tone (dial key off normal). The tone that reminds an operator to restore the dial key after a call has been completed into a step-by-step office, and after the called party has answered.

dial pulsing. The transmission of telephone address information by the momentary opening and closing of a dc circuit a specified number of times, corresponding to the decimal digit which is dialed. This is usually accomplished—as with the ordinary telephone dial—by manual operation of a finger wheel.

dial speed. The measurement, in number of pulses, that a rotary dial can transfer in a given amount of time. The dial speed of a typical rotary dial is 10 pulses per second.

dial tone. A 90 Hz signal (the difference between 350 Hz and 440 Hz) sent to an operator or subscriber indicating that the receiving end is ready to receive dial pulses.

disconnect signal. A signal (on-hook) from the calling and called subscribers, which notifies the operator or office switching equipment that the call is over and the connection should be released.

divided ringing. See ground return ringing.

dry circuit. A circuit over which voice signals are transmitted, and which carries no direct current.

DTMF (Dual Tone Multifrequency Signaling). A method of signaling in which a combination of two frequencies out of a possible eight are used to transmit numerical address information. The eight possible frequencies are 697 Hz, 770 Hz, 852 Hz, 941 Hz, 1209 Hz, 1336 Hz, 1477 Hz, and 1633 Hz.

duplex signaling (DX). A signaling system which occupies the same cable pair as the voice path, yet does not require filters. One duplex signaling section is confined to 5000 ohms of loop resistance, though several sections may be used in tandem.

E & M signaling. A signaling arrangement characterized by the use of separate paths for the signaling and the voice signals. The M lead (derived from "Mouth") transmits ground or battery to the distant end of the circuit, while incoming signals are received as either a grounded or open condition on the E (derived from "Ear") lead.

end office. The local central office at which subscriber lines and trunks are interconnected. It is designated a class 5 office in the DDD network.

E-type pulsing signal. In PCI (panel call indicator) pulsing, the end of pulsing signal informs the distant end that all digits have been sent. This signal is required when (PCI) traffic is completed through a crossbar tandem.

FSP (Frequency Shift Pulsing). A signaling technique which uses a frequency shift between 1070 Hz and 1270 Hz. This type of signaling is used with narrow band systems, such as teletypewriter switching networks.

full duplex. Telegraph or signaling circuits arranged for transmission in both directions at the same time.

go signal (start dial). A supervisory on-hook signal received by the calling end of a trunk after a stop signal has occurred. The go signal indicates that the called end is now ready to receive additional digits.

ground-return ringing (divided ringing). A party line system where ringers in a telephone circuit may be connected between one or both sides of the line and ground, with a capacitor placed in series with each ringer.

group-busy tone (all trunks busy tone). A low audible tone on the sleeve of trunk jacks at cord switchboards indicating that all trunks are busy. An absence of this tone informs the operator that there is at least one idle trunk in a group.

half-duplex. A communications system in which information can be

transmitted in either direction, but only in one direction at a time.

harmonic ringing. The technique of selectively signaling individual ringers on a party line using frequencies which are harmonics of fundamental frequencies. The fundamental frequencies used are 16-2/3 Hz and 25 Hz; the harmonics are 33-1/3 Hz, 50 Hz, and 66-2/3 Hz.

high-low signaling. Signaling in which a high resistance shunt indicates an on-hook condition, and a low resistance shunt an off-hook condition.

high tone. A −17 dBm, 480 Hz information signal that may, for example, be used as a partial dial tone.

hookswitch. A switch that is located within the supporting structure on which a telephone handset rests when it is not in use. When the handset is lifted, the switch closes the telephone circuit or loop.

howler tone. A 480-Hz tone which progresses in successive levels from 0 to 120 dBa, and is used to alert a subscriber when his handset is off-hook.

in-band signaling. The transmission of signaling information via tones at some frequency or frequencies that lie within a carrier channel normally used for voice transmission.

interoffice trunk. The telephone channel between two central offices.

intraoffice trunk. The trunk connection within the same central office.

keypulsing signal. The signal which indicates a circuit is ready for pulsing, in multifrequency and direct current keypulsing.

LAMA (Local Automatic Message Accounting). A combination of automatic message accounting equipment and automatic number identification equipment in the same office. In such a system, a subscriber-dialed toll call can be automatically processed without operator assistance.

leak and loop tests. Two tests designed to simulate the worst-case loop pulsing conditions in the exchange network. The tests are made at the central office switching equipment to assure that it will accurately respond to any pulsing signals generated by telephones connected to that exchange network.

local battery. A telephone circuit power source usually in the form of dry cells, and located at customer's end of the line.

loop. The closed circuit that is formed by the subscriber's telephone and the cable pair and other conductors that make the connection to central office switching equipment.

loop pulsing. Signaling achieved by the repeated opening and closing of the loop at the originating end of the circuit. Rotary telephone dials are loop pulsing devices.

loop signaling systems. Any of three types of signaling which transmit signaling information over the metallic loop formed by the trunk conductors and the terminating equipment bridges. Transmission of the loop signals may be accomplished by (1) opening and closing the dc path

around the loop, (2) reversing the voltage polarity, or (3) varying the value of the equipment resistance.

low-high signaling. A variation of high-low signaling. The on-hook condition is indicated by a low resistance shunt and an off-hook condition shows high resistance.

low-tone. A 480-Hz plus 620-Hz tone at −24 dBm. The 140-Hz difference (620-480=140) frequency gives the tone its low-pitched sound. A low tone is used for line busy, reorder, and no-circuit tone signals that are reached by a subscriber.

multifrequency pulsing. A method of transmitting address information (usually performed by an operator). The identity of each of the ten possible digits (0 to 9) plus the required supervisory functions is determined by a combination of two out of six possible frequencies.

no circuit (NC) signal (fast busy signal). A low tone (140 Hz) which is interrupted at 120 impulses per minute (IPM), and which indicates that there is no circuit available.

off-hook. An off-hook condition occurs when the telephone handset is lifted from its mounting, thus causing the hookswitch to operate (close), and closing the loop to the central office. The off-hook condition indicates a busy condition to incoming calls.

office alarm. An alerting signal indicating an abnormal condition in a central office.

office code. The first three digits of a seven-digit telephone number.

on-hook. An on-hook condition exists when the telephone handset is on its mounting, thus keeping the hookswitch open. The on-hook condition opens the dc loop, indicating that calls can be accepted.

ONI (Operator Number Identification). At a local dial central office, that equipment which allows the operator to come in long enough to acquire the calling number so that it may be keyed into CAMA equipment.

open-circuit signaling. Direct current signaling accomplished by opening and closing the telephone circuit. There is an absence of current flow when the circuit is in the idle condition.

out-of-band signaling. A method of signaling which uses a frequency that is within the passband of the transmission facility, but outside of a carrier channel normally used for voice transmission.

Panel Call Indicator (PCI) pulsing. A dc pulsing system where information digits are each transmitted as a series of four marginal, polarized pulses, which are detected at the receiving end and registered on relays or switches at terminating offices. In manual offices, received digits, as registered on the relays at the terminating office, are displayed in front of the operator.

partial dial tone. A high tone that notifies a calling party that he has not completed dialing within a specified period of time, or that not enough digits have been dialed.

percent break. The percentage of the total time of a pulse interval during which dial contacts or relays remain open.

polar relay. A permanent-magnet core relay which is designed to operate only when current flows in a specified direction.

polar signal. A signal whose information is transmitted by means of directional currents.

pulsating current. Current which varies in amplitude but does not change polarity.

pulse correction. The restoration in a dial pulse repeater of signaling pulses which have been distorted during transmission.

pulse link repeaters. A telephone repeater which connects one E & M signaling circuit directly to another E & M signaling circuit.

pulsing. The transmission of address information to a switching office by means of digital pulses. Pulsing methods include multifrequency, rotary dial, and revertive.

pulsing limits. The maximum amount of pulsing distortion that a central office can tolerate in the dial pulses generated by a customer's telephone, before the switching equipment begins to make errors.

register. (1) A device capable of storing digital information. (2) The first unit in the assembly of common control equipment in an automatic central office. The register receives address information in the form of dial pulses or dual tone multifrequency (DTMF) signals, and stores it for possible conversion or translation. A register frequently operates in conjunction with a sender.

restriction of message (toll diversion). A telephone arrangement where outgoing calls from a private automatic branch exchange (PABX) must either be routed through an operator or are limited to specified trunk groups.

reverse-battery signaling. A type of loop signaling in which battery and ground are reversed on the tip and ring of the loop to give an "off-hook" signal, when the called party answers.

revertive pulsing. Pulsing over a trunk in reverse direction. That is, from the terminating office instead of from the originating office. When the incoming office trunk is seized, it sends open or ground pulses to the office that originated the call. The originating office counts the pulses and opens the trunk when the correct number of pulses has been received.

ring. The ring-shaped contact of a plug usually positioned between, but insulated from, the tip and sleeve. The audible alerting signal on a telephone line.

ringback tone. An interrupted low tone indicating that the called telephone is ringing. Now called audible ringing tone.

ringdown. A type of signal that uses either a 135-Hz or 1000-Hz signal, interrupted 20 times per second. The type of signaling employed in manual operation, as differentiated from dial signaling. Ringdown signaling utilizes a continuous or pulsing ac signal transmitted over the line. The term ringdown originated in magneto telephone operation, where cranking the magneto of a subscriber set would "ring"

its bell and cause a marker to fall "down" at the central switchboard. In ringdown signaling, a key is operated in a cord circuit to ring on a trunk. On intertoll trunks, ringers are used to transmit and receive the signals. While ringdown trunks are unsuitable for intertoll dialing, connection of dial trunks to ringdown trunk can be provided with operator intertoll dialing.

ringing signal. Any ac or dc signal transmitted over a line or trunk for the purpose of alerting a party at the distant end of an incoming call. The signal may operate a visual or aural device.

ring trip. The circuitry required to disable the ringing signal when the called telephone is answered (placed in the off-hook condition).

SATT (Strowger Automatic Toll Ticketing). A system which, when a customer dials a toll call, automatically records a record of the calling number, the called number, the time of day, and the duration of the call.

selective ringing. A system designed with the capability of ringing only the desired subscriber's telephone on a multiparty line. Ringers tuned to one of five possible frequencies are used to achieve this effect.

semi-selective ringing. A four-party line ringing arrangement in which each party hears his own ring plus one other (usually a one and two ring code). Two of the parties on the line are rung from the tip of the line to ground, and the remaining two are rung from the ring of the line to ground.

sender. A unit which receives address information from a register or routing information from a translater, and then outpulses the proper routing digits to a trunk or to local equipment. Sender and register functions are often combined in a single unit.

separate-channel signaling (a type of CCSS). A carrier system signaling arrangement where the signaling for several channels is multiplexed on a single voice channel.

signaling. The process by which a caller on the transmitting end of a line informs a particular party at the receiving end that a message is to be communicated. Signaling is also that supervisory information that lets the caller know that the called party is ready to talk, that his line is busy, or that he has hung up. Signaling also holds the voice path together while a conversation goes on.

simplex signaling (SX). Signaling over a trunk circuit. Signaling information is sent over the circuit when a signal-transmitting relay at the calling end is energized. A signal-receiving relay at the called end provides the signaling information to the voice channel. Both relays are connected to the midpoints of repeating coils or retardation coils at each end of the circuit.

single-frequency signaling. A method of signaling in which a single frequency tone, 2600 Hz for example, is placed on the voice path. The tone is on during the idle condition, pulsed during dialing, and off when the circuit is being used. (This condition is known as tone-on-when-idle.)

sleeve (S). (1) The third contacting part on a telephone plug—preceded in location by the tip and ring. (2) The

sleeve wire is the third control wire of each telephone in an automatic switching office.

speech-simulated signal. A signal made up of those components of a voice signal which will cause the false operation of tone-operated supervisory equipment.

start-dialing (start pulsing) signal. The on-hook condition indicating that the receiving end is ready to receive pulsing information.

step-by-step. (1) The process by which a call is progressively carried to the desired terminal under the direct control of subscriber-initiated pulses, or pulses from a sender. (2) Also refers to an electromechanical rotary switching system.

subscriber's loop. See "loop."

subset. A subscriber's telephone apparatus.

superimposed ringing. A type of semi-selective ringing which uses a combination of ac and dc.

supervisory signal. A signal which indicates whether a circuit is in use. A signal which gives an indication of status or change of status in a telephone system. For example, a signal used to attract the attention of an operator.

switching office. A location where either toll or local telephone traffic is switched or connected from one line or circuit to another. Also called switching center.

synchromonic ringing. A type of party line selective ringing which uses ringing frequencies of 16 Hz, 30 Hz, 42 Hz, 54 Hz, and 66 Hz.

talking battery. The dc voltage supplied by the central office to the subscriber's loop to operate the carbon transmitter in the handset.

talking path. In a telephone circuit, the transmission path consisting of the tip and ring conductors.

TASI (Time Assignment Speech Interpolation). Because audible speech on the average voice circuit is only present approximately 45% of the time, the efficiency of the expensive overseas channels can be improved if some of the remaining 55% of the time could be utilized. TASI switching equipment connects a party to an idle circuit while speech is taking place, but disconnects the party when speech stops, so that a different party can use the same circuit. During periods of heavy traffic, TASI improves line efficiency from 45% to 75-80%.

through supervision. Total supervision of a toll call by the originating operator through any intermediate switching points to the called telephone.

time-division signaling. Signaling over a time division multiplex system in which all voice channels share a common signaling channel, with time division providing the separation between signaling channels.

tip. The contacting part at the end of a telephone plug or the top spring of a jack. The conductors associated with these contacts.

tone signaling. The transmission of supervisory, address, and alerting signals over a telephone circuit by means

of voice frequency tones. Also used in "Touch Tone" dialing.

trunk. (1) A communications channel connecting two switching centers, or a switching center with an individual terminal. (2) A communications channel between two offices or between equipment within the same office. A trunk is used commonly for all calls of the same class that are generated between two terminals. (3) The different types of trunks, some of which can be classified as (a) direct trunk—interconnects two class five end offices, (b) toll-connecting trunk—connects a class five end office to any higher ranking toll office, (c) intertoll trunk—connects any class one through four toll switching office to any other class one through four office.

two-state signaling. Any signaling system which only transmits two states, such as "on hook" and "off hook", or "off" and "on". Such systems are often used for transmission of dial pulses or supervisory signals.

wet circuit. A circuit which carries direct current.

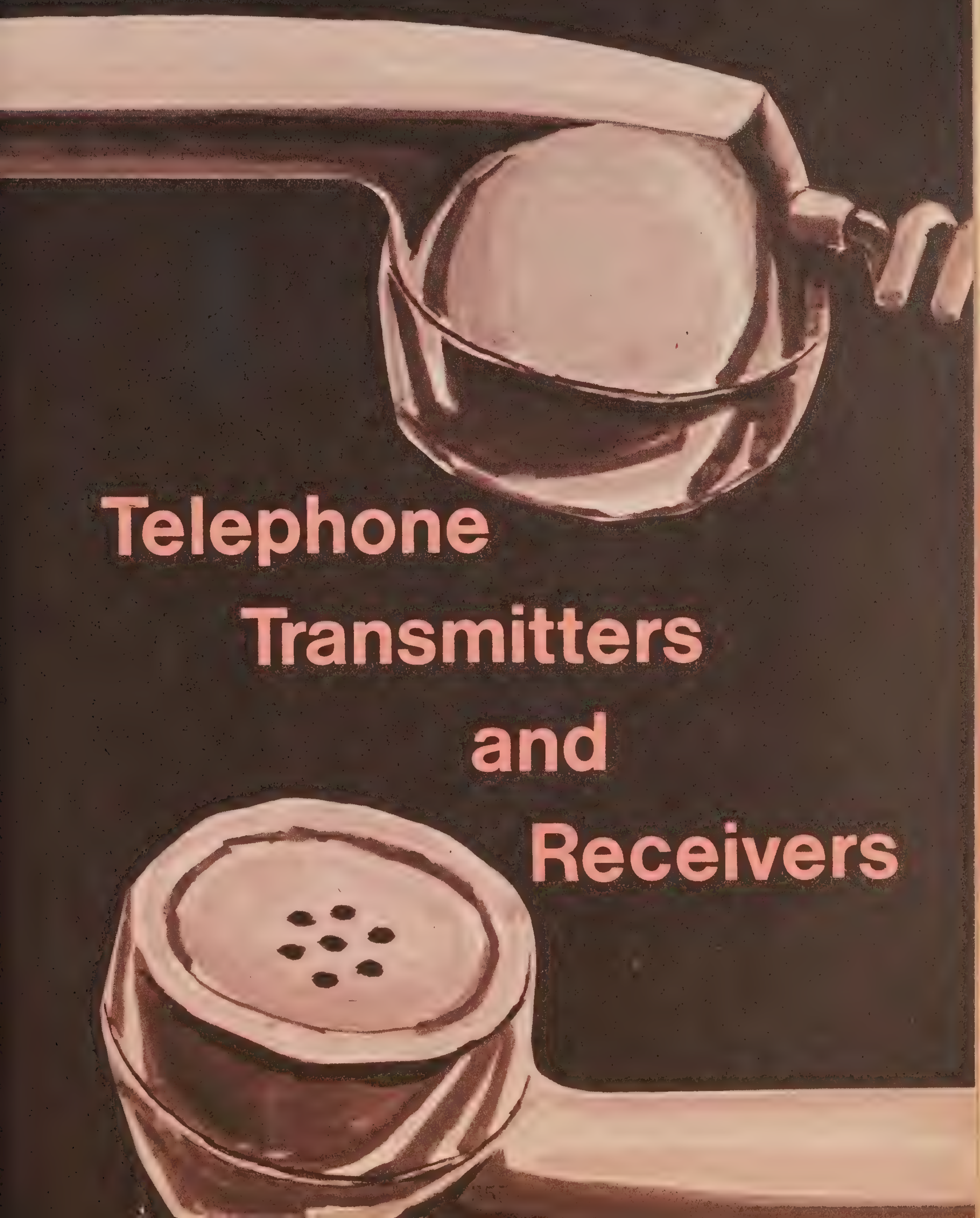
wink operation. In wink operation, trunk equipment sends on-hook signals toward each end during the idle condition. When a connect signal is received, the called office requests a register or sender. However, the on-hook signal to the calling office remains until the register or sender is connected at the called office. At that time, the idle on-hook signal changes to off-hook. The register or sender retains the off-hook signal for not less than 140 milliseconds, then returns an on-hook signal to the calling end, indicating that it is ready to receive pulses from the calling office.

wink signal. A short interruption of current to a switchboard busy lamp to indicate that the circuit is busy. On key telephone sets, the wink signal indicates that a line is being held. It is also an indication of change of state between an on-hook and off-hook condition.

GTE LENKURT

DEMODULATOR

MAY 1974



Telephone Transmitters and Receivers

The word "telephone" literally means "sound at a distance." While telephone equipment has changed considerably over the years, the basic principles that made "sound at a distance" possible remain the same.

In telecommunications, the basic goal is the transmission of intelligence with a minimum of distortion, from the point of origin to the intended destination. This intelligence may be in the form of speech, a scene, a picture, a series of electrical signals, or a modulated electric wave. A telephone system establishes a voice communication path between any two telephones in that system. To do this successfully, the system must convert sound waves to electrical signals and back again in such a manner that the distant receiver hears the original voice sounds with a minimum of distortion. The telephone system must also establish a communication channel and transmit information over it as quickly and reliably as possible.

The difference in nature between audible sound signals and electromagnetic signals necessitates the utilization of devices or converters designed to permit passage from one medium to the other. While in theory, and to a certain extent in practice, it is possible to produce one device that converts sound to electromagnetic impulses and back, power limitations have forced the design of two separate converters. The converter that changes sound variations to corresponding electromagnetic variations is called a transmitter. The converter that changes electromagnetic variations back to corresponding sound variations is called a receiver. (See Figure 1.)

Telephone Transmitters

A telephone transmitter converts mechanical vibrations in the air to electrical "sound" vibrations in an electrical circuit. This is accomplished by varying the resistance in a portion of the transmitter in step with the incoming sound. The varying resistance causes a direct current to increase or decrease.

The simplified telephone transmitter shown in Figure 2 consists essentially of two components—a diaphragm and a capsule filled with carbon granules. The diaphragm is usually made of a stiff metallic material, and is held in place by a retaining ring. A rod connects the diaphragm to the car-

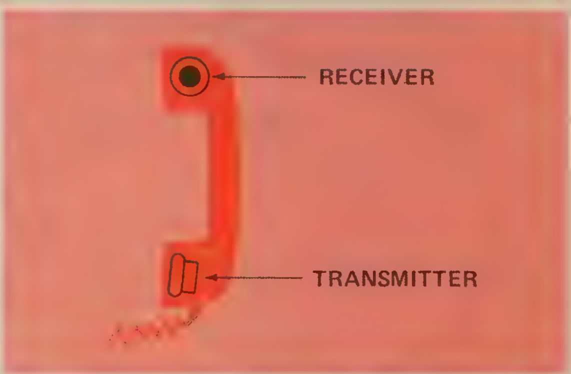


Figure 1. The transmitter and receiver in the telephone handset convert audible sound signals to electromagnetic signals, and back to audible sound.

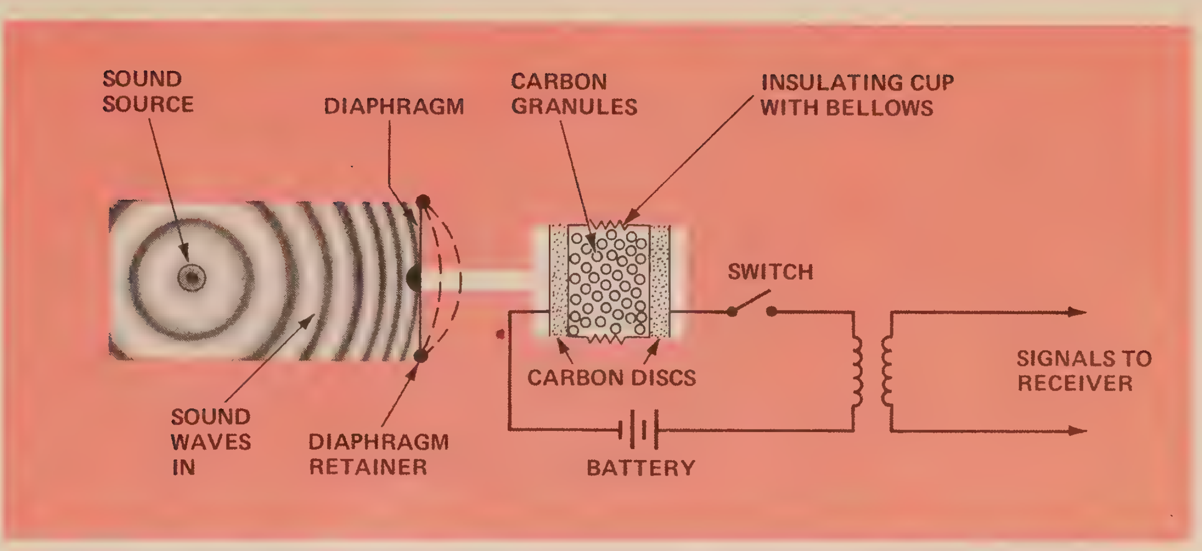


Figure 2. A simplified telephone transmitter.

bon-filled capsule. The capsule is made up of two metal-backed carbon discs held together by an insulating cup with paper bellows. The space between the discs is filled with carbon granules. Because of the action of the bellows, one disc moves freely with the diaphragm. The other disc is rigidly held. The whole capsule is part of a circuit through which direct current can flow.

The components of the transmitter change sound signals to electrical signals. Sound waves meeting the diaphragm cause it to move back and forth. These movements are transmitted to the movable carbon disc. The carbon granules between the two discs are then alternately compressed or decompressed in step with the incoming sound. The compressions and decompressions (rarefactions) vary the resistance of the granules causing the direct current (dc) in the circuit to increase or decrease. The mechanics of the resistance changes are shown in Figures 3 and 4. (While carbon granules are irregular in shape, they are shown here as round for simplicity.)

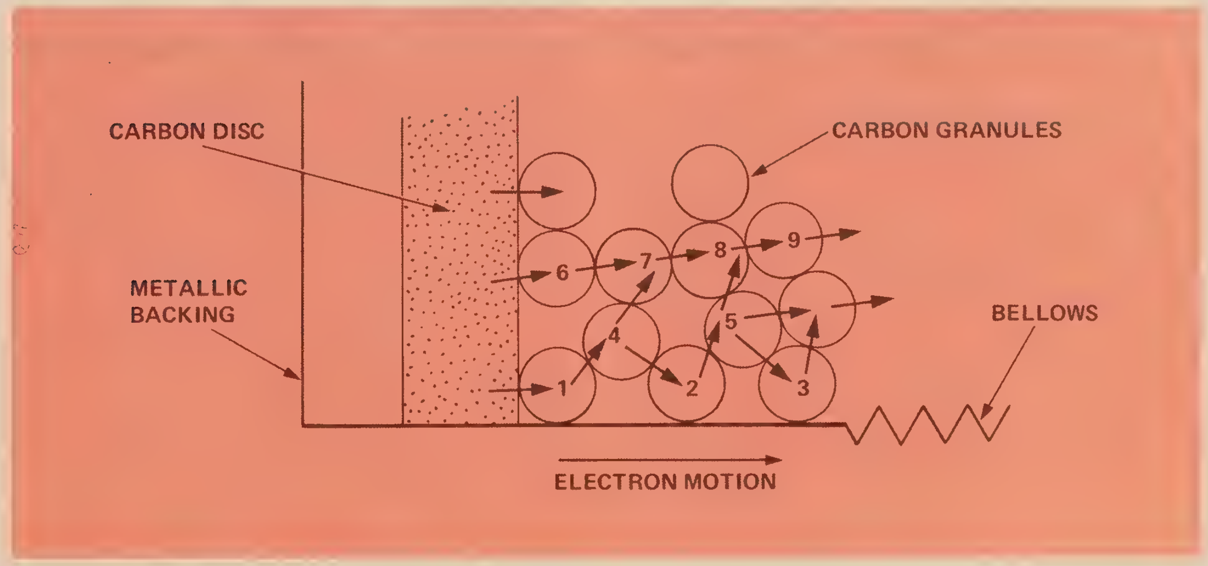


Figure 3. Carbon granules in a normal or decompressed condition within a telephone transmitter capsule.

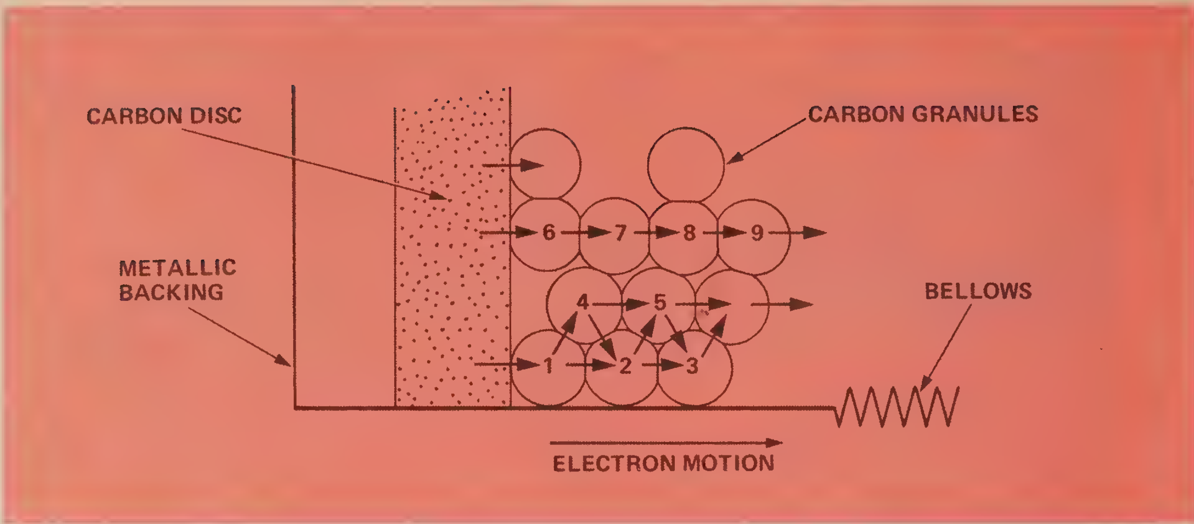


Figure 4. Carbon granules during a period of compression.

Figure 3 represents a steady-state condition with no sound meeting the diaphragm, and a steady current in the circuit. There is a possibility of several paths for the current. It can, for example, travel from the carbon disc to granule 1, 4, 2, 5, 3, etc. Figure 4 shows an instant in time when the granules are compressed by a wave. Granules 1, 2, and 3 have now come in contact, thus providing an additional path for the current. Also, granules 6, 7, 8, and 9 have been compressed so that instead of there merely existing point contacts, there are circular-area contacts between them. This increase in contact area decreases the overall resistance and thus increases the current. The high resistance caused by a decompression wave is produced when the carbon disc moves to the left (see Figure 4), allowing the granules to separate; in some cases contacts are broken and in other cases only point contacts are established. The undulating voice current (see Figure 5) produced by the compression and decompression of the carbon granules is carried by the line to the distant end where it is coupled to a receiver circuit and changed back into sound.

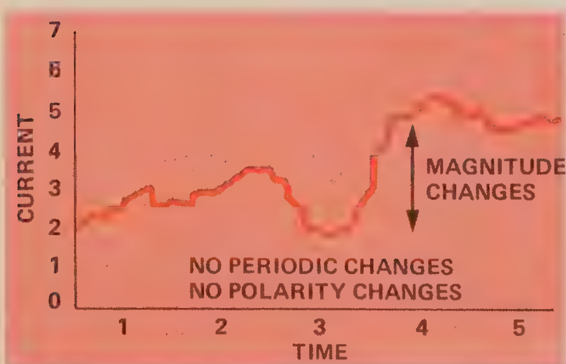


Figure 5. An undulating voice current is produced by the compression and decompression of the carbon granules in the telephone transmitter.

Telephone Receivers

A telephone receiver changes electric voice signals to audible sound by magnetically vibrating a diaphragm in step with the incoming electric voice signals. Voice signals enter the receiver circuit over a coupling coil. The signals continue to the receiver capsule, where they pass through coil windings and generate a fluctuating magnetic field. The magnetic field causes a rigid diaphragm to vibrate, thus producing an audible sound.

There are two types of receivers which are most commonly used—the biased or direct-action receiver and the polar or indirect-action receiver. In the

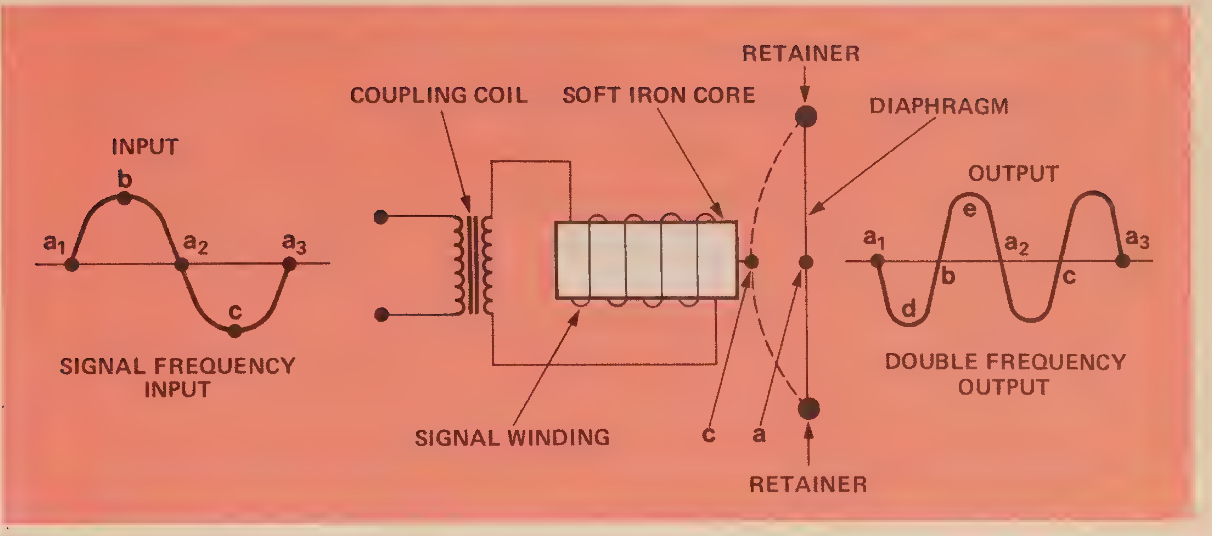


Figure 6. A receiver with no permanent magnet incurs double-frequency distortion problems.

first type, a magnetic diaphragm is directly vibrated by the fluctuating field. In the second type, the field vibrates a magnetic armature which then transmits the vibrations through a linkage to a non-magnetic diaphragm.

Direct-Action Receiver

The direct action or biased receiver has been a standard unit for many years. Its diaphragm is the link between the electric circuit and the air. The voice signals are changed into magnetic field fluctuations by two coils. These fluctuations then vibrate the magnetic diaphragm. The principal parts of the direct-action receiver are two inductive windings, the diaphragm, and a permanent magnet. The permanent magnet is used to prevent double-frequency distortion.

An example of the cause of double-frequency distortion is shown in Figure 6. The receiver in Figure 6 has no permanent magnet, the signal winding is wound around a soft iron core which becomes magnetized only while current is flowing in the winding. When a signal is introduced, the core will be magnetized during each half-cycle. The diaphragm will be attracted as the signal increases from a_1 to b; then as the signal decreases from b to a_2, the diaphragm will be released. When the diaphragm is attracted, the air experiences a decompression; when it is released, the air is compressed. Thus, while the input signal goes through a half-cycle (from zero at a_1, to maximum positive at b, and back to zero at a_2), the output goes through a complete cycle (from zero at a_1, to maximum negative or decompression at d, to zero at b, to maximum positive or compression at e, and back to zero at a_2). The series of events described above are repeated on the negative half-cycle of the input. Each input cycle, therefore produces two cycles in the output, thus creating double-frequency distortion.

The direct-action receiver shown in Figure 7 was designed to prevent double-frequency distortion. The permanent magnet biases the diaphragm in an operating region where double-frequency distortion is at a minimum. The signal windings are wound on the permanent magnet poles in such a way that a positive going pulse will weaken the poles, and a negative going pulse will strengthen the poles. When the

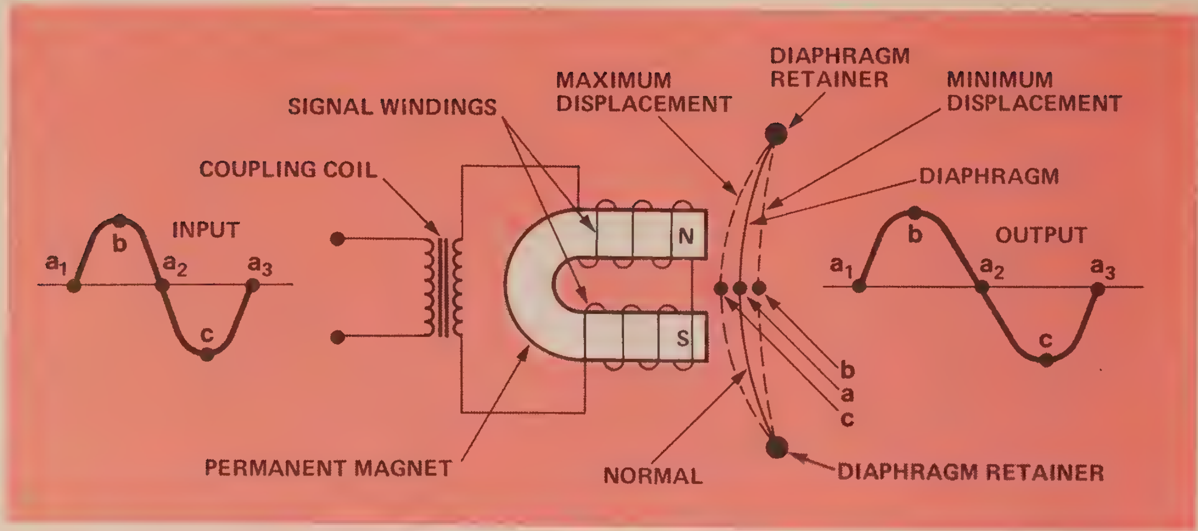


Figure 7. The use of a permanent magnet in a biased or direct-action telephone receiver prevents double frequency distortion.

input rises from a_1 to a maximum positive at b, and back to a_2, the diaphragm is released and returned, producing a compression. On the negative swing of the input signal (from a_2 to maximum negative at c, and back to a_3), the diaphragm is attracted and returned to normal, producing a decompression. This action generates an output resembling the input. The same results may be obtained by introducing a direct current into the receiver of Figure 6. However, there is a certain amount of amplitude distortion. On the negative half-cycles of the input, the diaphragm offers greater elastic resistance than on the positive half-cycles. This causes the decompressions to be weaker in amplitude than the compressions.

Indirect-Action Receiver

The indirect-action or polar receiver (shown in Figure 8) is a more recent development than the direct-action receiver; it is more sensitive and introduces considerably less distortion than the direct-action receiver. In effect, the indirect-action polar receiver is capable of discriminating the polarity

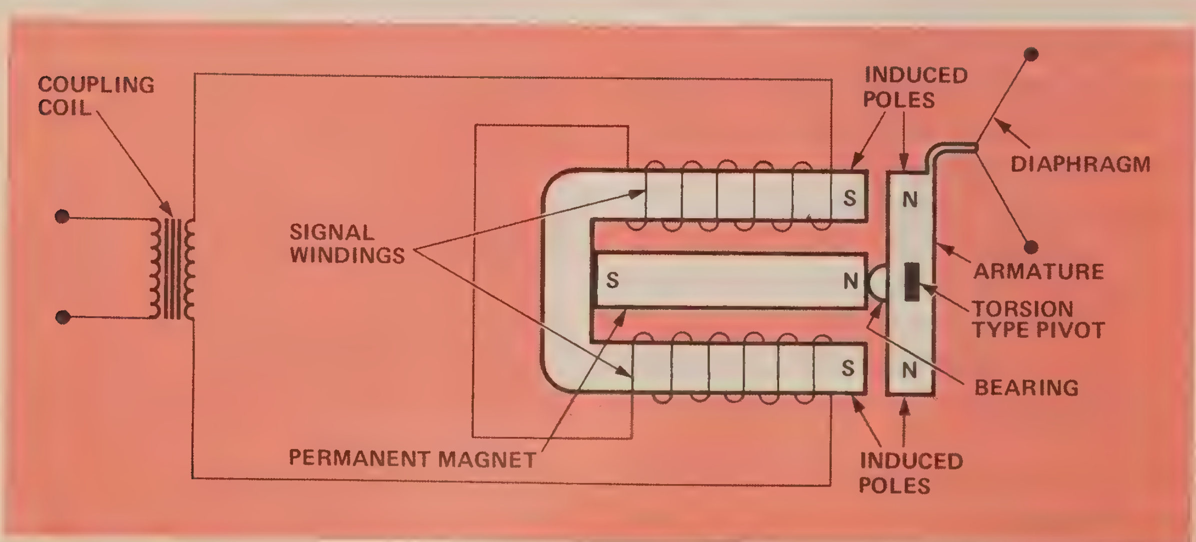


Figure 8. A polar or indirect-action telephone receiver.

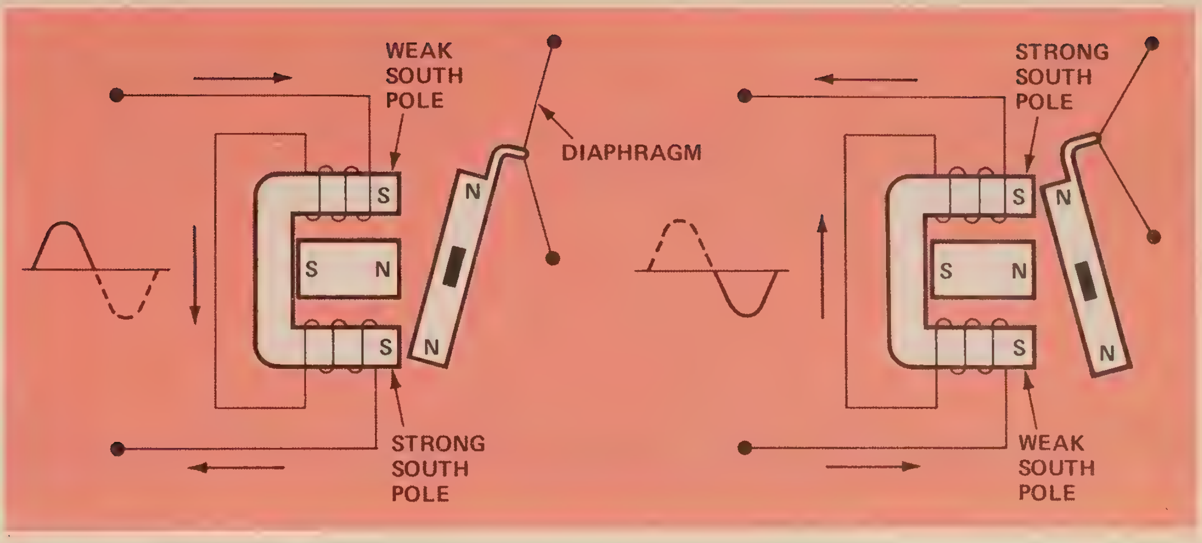


Figure 9. Polar receiver action.

of the incoming signals. The principal parts of this receiver are two signal windings, a permanent magnet, a movable armature, and a diaphragm.

The permanent magnet in Figure 8 polarizes the movable armature and the signal winding cores as indicated. The armature is held in normal position by a torsion bar and the equal attraction of the winding cores. The windings are wound so that current in one direction produces a magnetic field that aids the induced pole in one core and opposes the induced pole of the other. When the current reverses, the above effect also reverses. These unequal forces cause the armature to see-saw about its pivot and in this way, the diaphragm, which is attached to one end of the armature, moves back and forth as shown in Figure 9. One advantage of the indirect-action receiver is that the diaphragm is designed specifically for producing sound and does not have to serve the dual purpose of being a good magnetic conductor as well as a good acoustical generator. Also, less electrical energy is required because the magnetic circuit is more compact and efficient.

The various components of a telephone system can be overwhelming if one attempts to perceive them all at once; but taken individually, it becomes much easier to visualize an overall system. This is especially true today, when the business of telephony encompasses not only basic telephone principles and associated apparatus, but also other forms of telecommunications, such as video and machine communications.

BIBLIOGRAPHY

1. *Electrical Principles of Telephony, General System Practices (Section 945-050-099).* Northlake, Illinois: GTE Automatic Electric Incorporated, 1962.

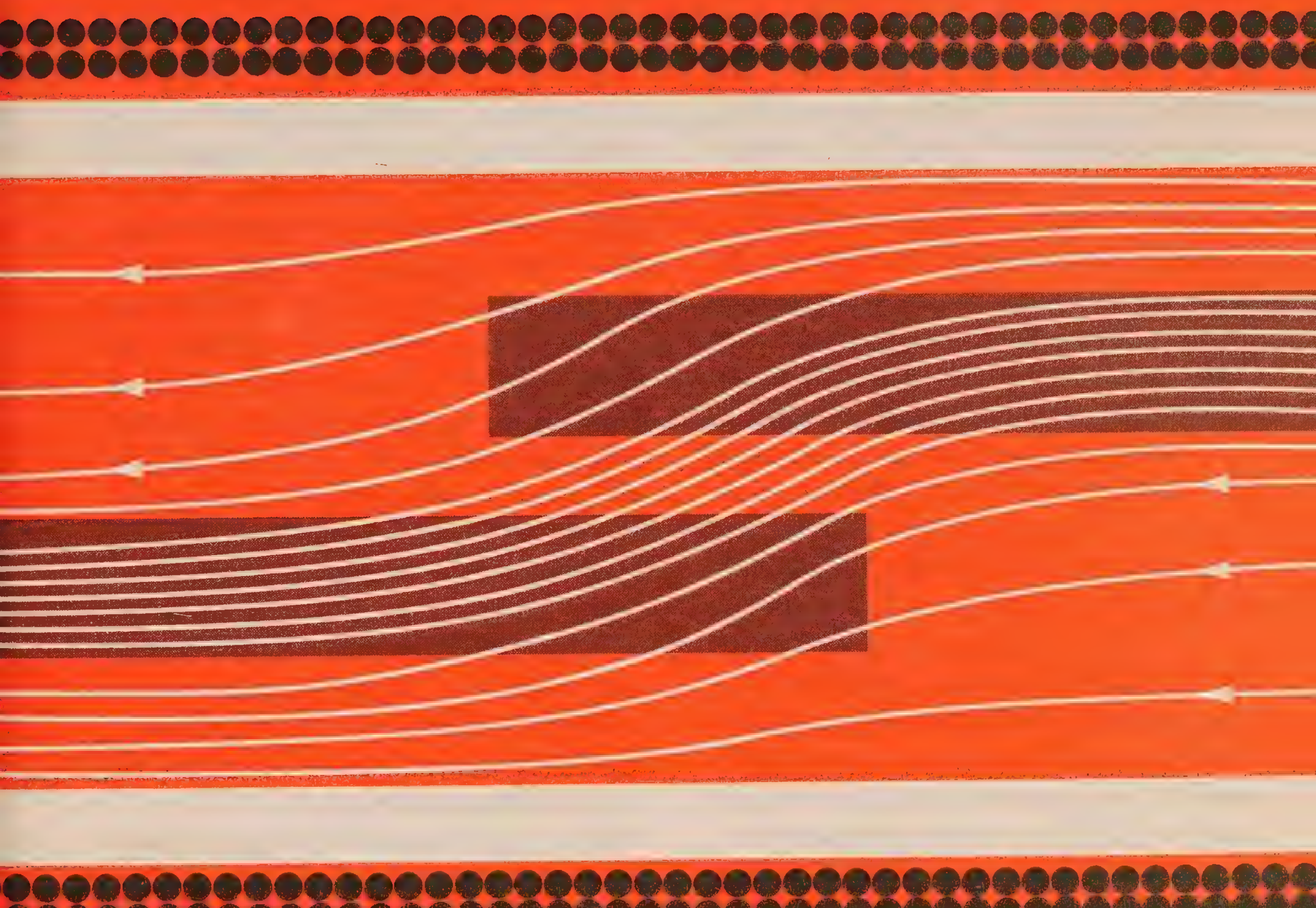
2. *Telephone Principles and Basic Apparatus, General System Practices (Section 945-100-010).* Northlake, Illinois: GTE Automatic Electric Incorporated, 1964.

3. *Principles of Electricity Applied to Telephone and Telegraph Work*, American Telephone and Telegraph Company, June 1961.

GTE LENKURT

DEMODULATOR

JUNE 1974



Basics of Telephone Relays

The relay is one of the first electromagnetic devices developed and used for communications. It was an essential part of the old telegraph networks, and its name is derived from the function it performed in those networks.

In the early days of electrical communication, attempts to transmit signals over longer and longer distances met with little success. The dots and dashes of Morse Code became too weak to operate any kind of sounding device, after being transmitted over miles of wire. The problem could be corrected by increasing the power of the transmitted signals, but this was only a partial solution and ultimately proved to be impractical.

Figure 1 shows a direct telegraph communication circuit between two stations. If stations A and B are connected by a line whose loss characteristics are such that only 1/1000th of the original signal power transmitted from A arrives at B, and the sounding device at B requires a minimum of 1 watt to operate, then A will have to send a 1000 watt signal to be heard at B. If it is later necessary to expand the network by installing station C (see Figure 2), and the line connecting C to B is of the same length and characteristics as the line connecting B and A, and the sounding device at C also has a 1 watt power requirement, then for B to communicate with C, 1000 watts of power is necessary. For A to directly communicate with C, 1,000,000 watts is required, since to receive the necessary 1 watt at C there would have to be 1,000 watts at B. The total requirement would thus be 1,000,000 watts from A. The use of such large quantities of power could be avoided by having station A send a message at 1000 watts, and have an operator at B send the message on to C at an additional 1000 watts. This solution, however, still requires excessive power. In addition, the human repetition involved is a source of error and introduces a definite delay factor. The development of the relay produced a device that was used to great advantage in solving the early problems of electrical communications.

The relay is an electromagnetically operated switch used to open and close electric circuits. It changes electric action to mechanical action. The essential parts of a relay are shown in Figure 3. When current flows through a coil wound on a soft iron core, the

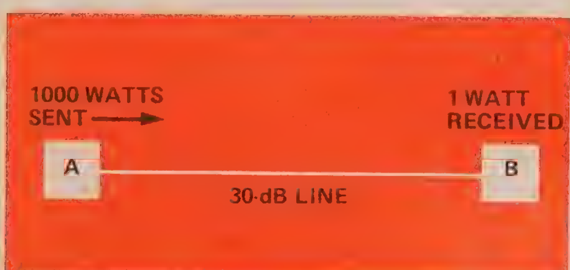


Figure 1. A direct telegraph communication circuit between two stations.

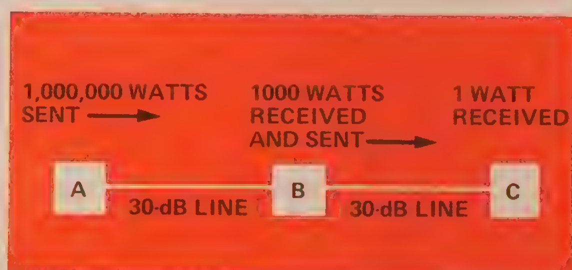


Figure 2. The expansion of a telegraph line necessitates an increase in power.

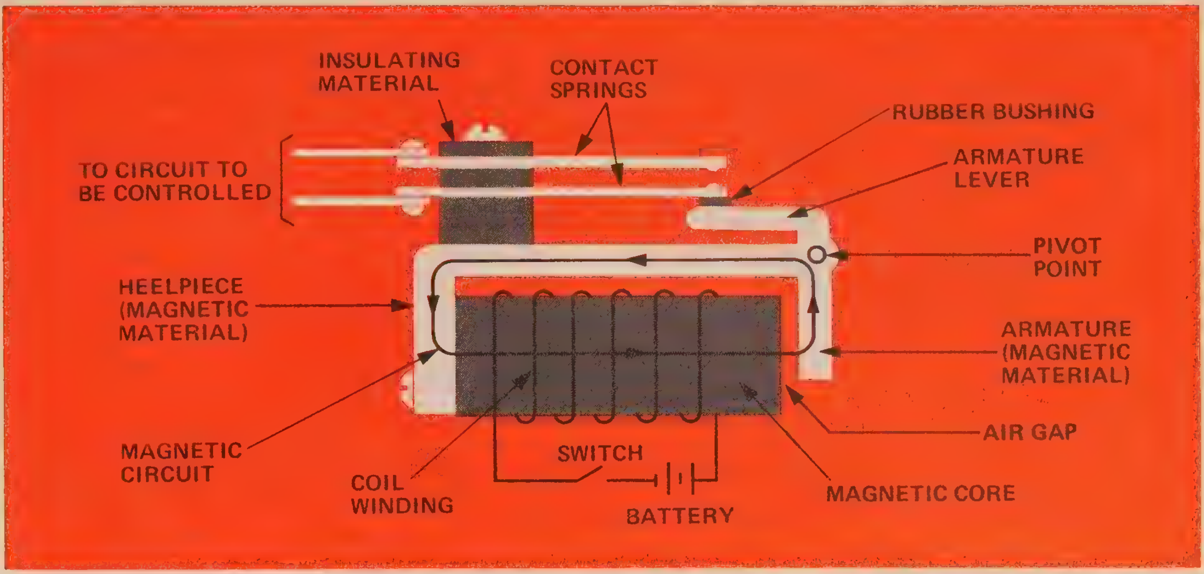


Figure 3. Current flowing through a relay coil magnetizes the core, which then attracts the relay armature.

core becomes magnetized. The magnetizing action of a weak current may be increased by increasing the number of turns on the core. For example, a relay coil may be wound so that it will operate with one milliwatt (1/1000th of a watt) of power. The core of the relay may then generate a magnetic field that is strong enough to attract the movable armature, the motion of which causes the contact springs to close. This closing of the springs may then be used to complete a circuit that can generate larger amounts of power. When the magnetizing current is removed, the core is demagnetized, and the armature is returned to its normal position by the spring action of the contact springs.

By utilizing relays in the hypothetical communication network of Figure 4, not only can the power requirement be reduced, but the human repetition can also be eliminated at the relay point. If this relay is capable of operating on 1 milliwatt, telegraph messages are sent from station A to station B on only 1 watt of power. If station A wishes to communicate directly with station C, a switch is provided as shown in Figure 4, and the relay at B controls a 1 watt sender instead of the telegraph sounder (a relay-type device that produces audible clicks when its armature moves from one position to another). To send messages from A to B, only 2 watts are required, one watt for transmission and one watt to operate the sounder. Communication between A and C requires 3 watts.

Modern Relay Uses

Today, relays are used primarily as switches and control devices. In this capacity they perform various functions that are required in the orderly and logical operation of complex communication networks. The relay is one of the most important components used in telephone switching, since nearly all telephone circuits depend upon one or more relays for operation. Relays also perform an important role in the remote control of certain operations, and in controlling the time of operation such as when:

(a) a weak or variable signal must have positive control over a local circuit,
(b) one circuit must cause changes in several other circuits,
(c) it is necessary to effect time delays or sequential operations or both,
(d) events which must occur simultaneously or in a certain sequence must change another circuit.

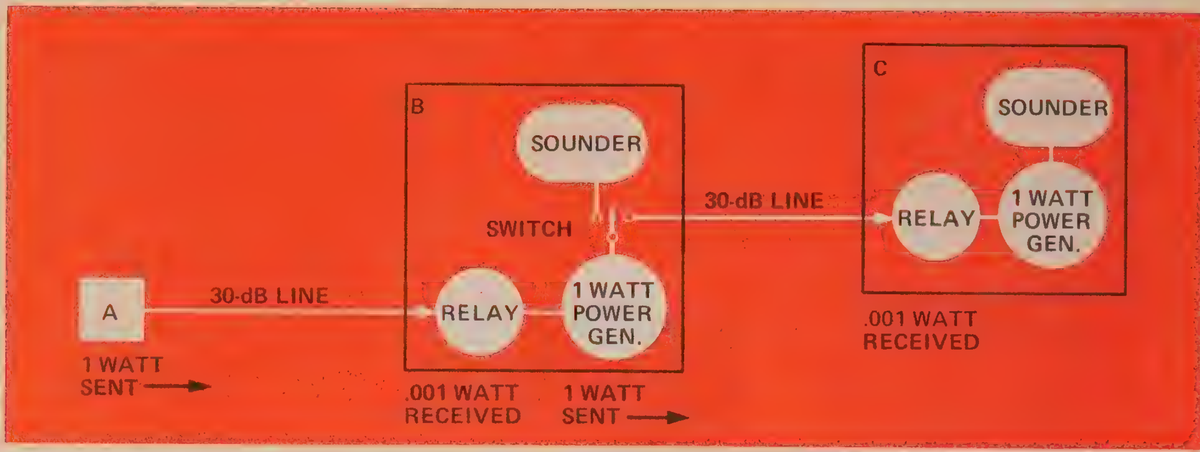


Figure 4. The use of relays in a communication network reduces power requirements as well as the possibility of human error.

Slow-Acting Relays

In today's automatic switching circuits, slow-acting relays are used to momentarily delay certain circuit operations. Slow-acting relays may be separated into two categories: the slow-operating relay and the slow-releasing relay. The slow action of these relays is accomplished by the use of a copper collar mounted around the coil core of the relay, which momentarily delays the operation of the relay armature (see Figure 5). The relay is made to both close and release slowly when the collar is on the armature end of the core. A slow-release-only action takes place when the collar is located on the heel end of the core. A copper sleeve around and along the entire core length also makes a relay slow to operate, and very slow to release. Slow-acting relays are mainly used when it is necessary that the relay contacts remain closed while the coil circuits are repeatedly interrupted for short periods.

Slow-Operating Relay

The slow operating relay is slow to attract its armature after the control circuit has been closed. The copper collar mounted on the armature end of the relay (see Figure 6) causes a delay in the attraction of the armature. To more easily visualize the action that takes place, the relay winding might be considered the primary of a transformer and the copper sleeve a short circuited secondary winding consisting of a single turn having a very low resistance. When a voltage is initially applied to the terminals of the winding, the current tends to build up and establish a magnetic field in the relay core. The instant the lines of force cut through the copper collar, a voltage is induced, causing a current to flow in the collar in the opposite direction from that in the winding. The current in the copper collar sets up a field in the same magnetic path, which opposes the field being built up by the current in the relay winding. Gradually the field in the copper collar dies away and the magnetism due to the winding builds up until it reaches a maximum value and attracts the relay armature.

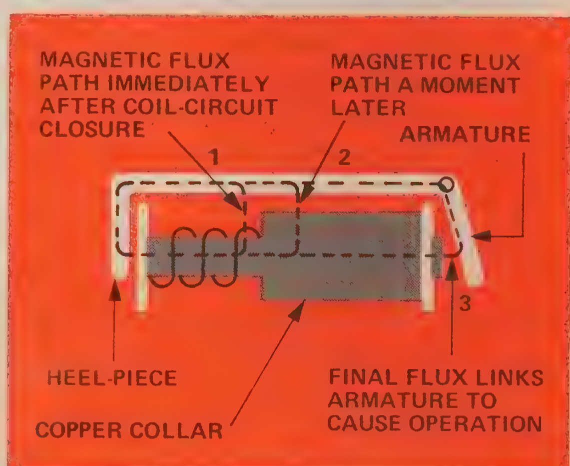


Figure 5. Magnetic flux paths in an armature-end collar relay.

Slow-Releasing Relay

The slow-releasing relay holds its armature momentarily, after the control circuit to the relay has been opened. The copper collar of the slow-releasing relay is mounted on the heel end of the relay (see Figure 7). When the circuit to the coil is broken, the magnetic field, in collapsing, sets up a current in the copper collar, which flows in a direction that attempts to maintain the existing magnetism. But, since this current itself is dependent upon the decreasing magnetism, both the field in the copper collar and in the relay, gradually die away. This action delays the release of the armature.

Correed Relay

An advancement beyond that of the conventional relay came with the development of the correed relay. The complete correed relay assembly consists of a core on which the operating coil is wound, and within which are one or more "reedcapsules." The reedcapsule is a hermetically sealed glass tube containing two flat ferromagnetic metal reeds, each supported at the ends and overlapping at the center of the capsule (see Figure 8). When the operating coil is energized, the magnetic field tries to follow the path of least resistance. The field will assume the

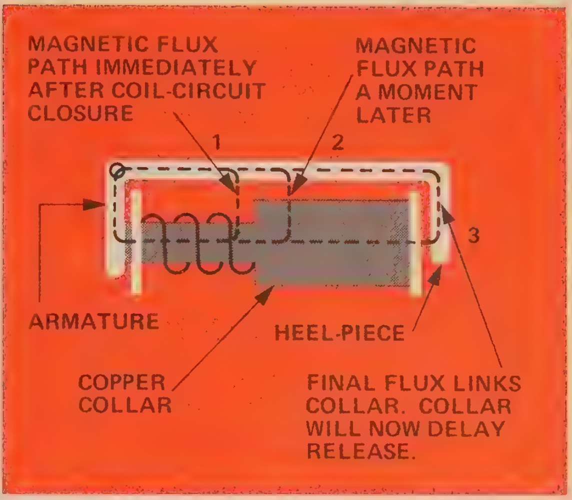


Figure 6. Magnetic flux paths in a heel-end collar relay.

shape and proportional distribution shown in Figure 9. The contact points magnetize with the indicated polarity and attract each other. When the coil is de-energized, spring action causes the reeds to separate. Some of the advantages of correed relays are lightness of weight, economy, and dependability. They may be constructed to contain several sets of contacts, as do other types of relays. Because of their compactness and excellent operating characteristics, correed relays find numerous applications in many of today's electronic switching systems.

Telephone Ringers

Telephone ringers usually operate in conjunction with some type of relay action. Ringers announce incoming

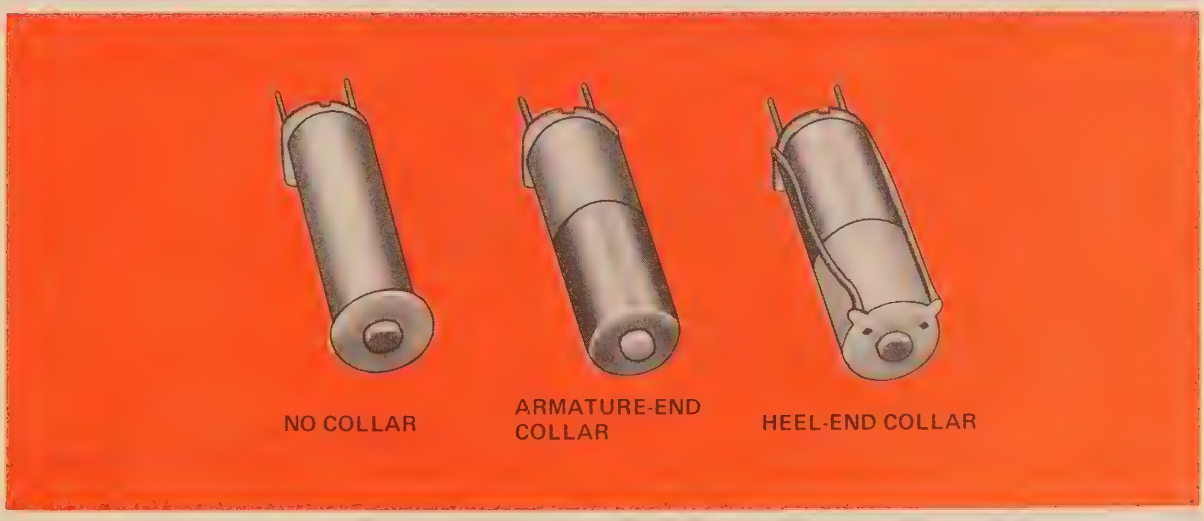


Figure 7. Typical relay coils.

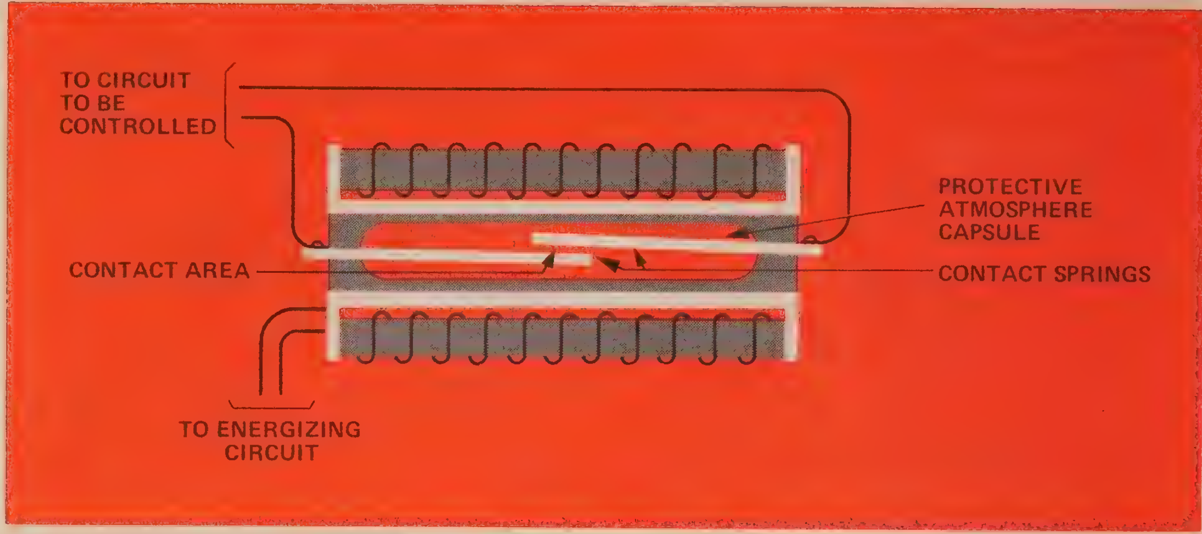


Figure 8. A correed relay.

calls. The working parts of a telephone ringer are shown in Figure 10; they consist of two electromagnetic coils, a permanent magnet, a pivoted armature with hammer, and two gongs. Except for a difference in size and physical arrangement, the operating principle of the ringer in Figure 10 is the same as the operating principle of the polarized telephone receiver (see the May, 1974 Demodulator). With no current in the ringing coils, the ends of the armature are of one polarity, and the ends of the coil cores are of the opposite polarity. Because of the equal attraction of the armature ends by the two coil cores, a relatively small bias spring force will hold the armature against one pole. With the ringing current (ac) applied to the coils, the magnetic balance is disturbed. Each half cycle causes first one, then the other electromagnet to attract its respective end of the armature with greater force. When the right-hand core shown in Figure 10 attracts the right-hand end of the armature, no action takes place; the armature remains in the position where it is normally held by the bias spring. When the left-hand core attracts the left-hand end of the armature, the bias spring force is overcome. The armature swings and the hammer (clapper) strikes the right hand gong. At the end of the half-cycle, the armature swings back to its normal neutral position, and the whip of the clapper rod causes

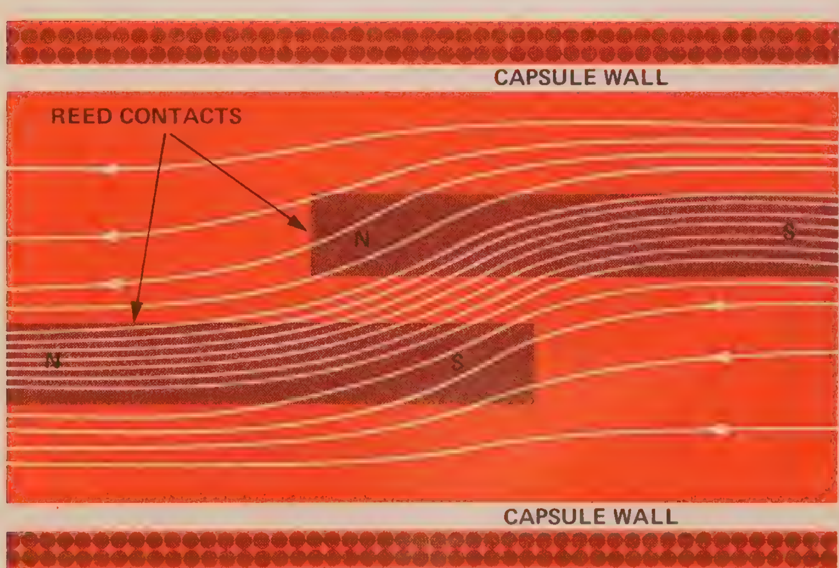


Figure 9. Magnetic lines of force in a correed relay.

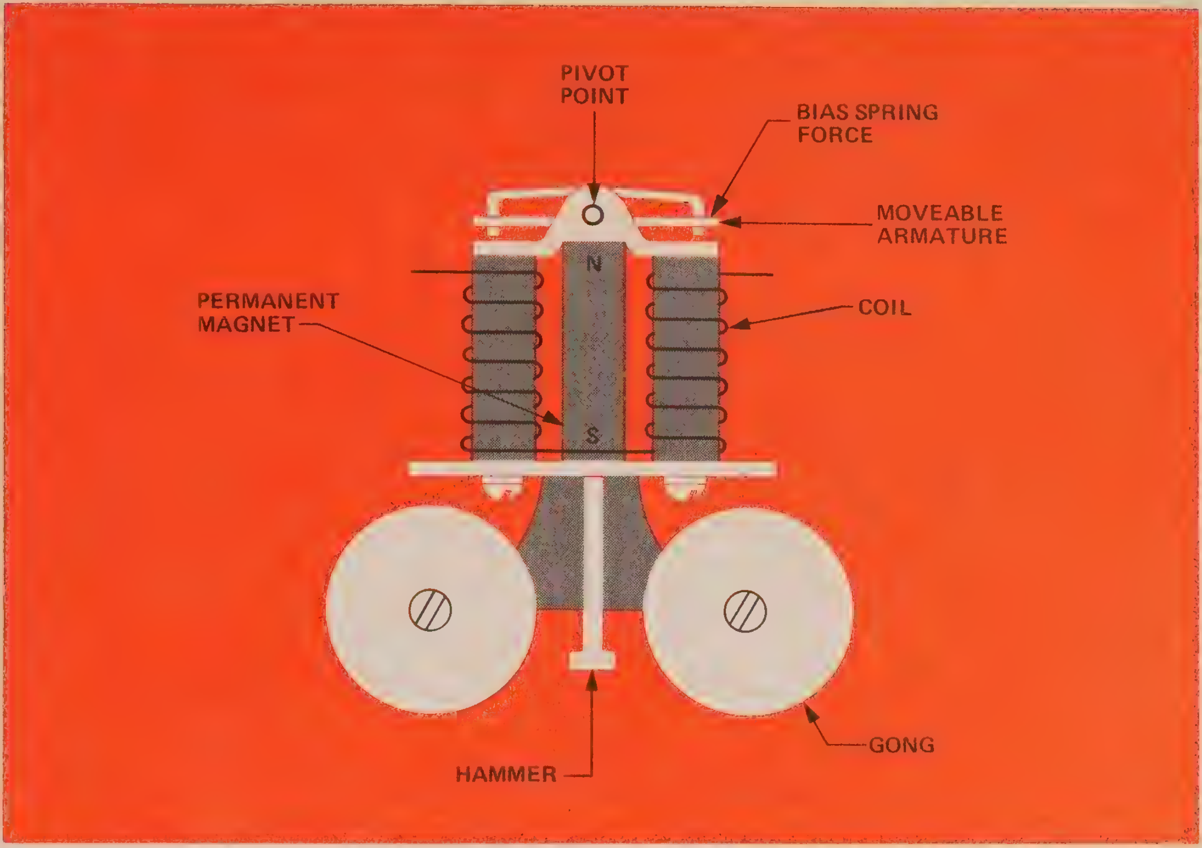


Figure 10. Typical telephone ringer arrangement.

the clapper to strike the left-hand gong. Since a biased ringer is not affected by one polarity of current in its coils, it can be connected to prevent bell tapping during dialing, or used with a cold cathode tube for full selective superimposed ringing.

A relay is essentially a switch that is operated by an electromagnet. The uses for this device run the gamut from simple local switching of a large current, to selecting and operating complex functions from a central point many miles away. In recent years, high-speed, solid-state devices have come into use in switching circuits where relays formerly dominated. However, where high-speed operation is not essential, the relay is often the choice, not only because of economy, but also because relay contacts and coils can better withstand those voltages and currents inherent in customer lines and trunks.

BIBLIOGRAPHY

Electrical Principles of Telephony, General Systems Practices (Section 945-050-99). Northlake, Illinois: GTE Automatic Electric Incorporated, 1962.

Goeller, L.F. Jr., "Telephone Switching Methods and Equipments," in D. Hamsher, ed. *Communications System Engineering Handbook.* San Francisco: McGraw-Hill Book Company, 1967.

Mechanical Principles of Telephony, General Systems Practices (Section 945-050-100). Northlake, Illinois: GTE Automatic Electric Incorporated, 1966.

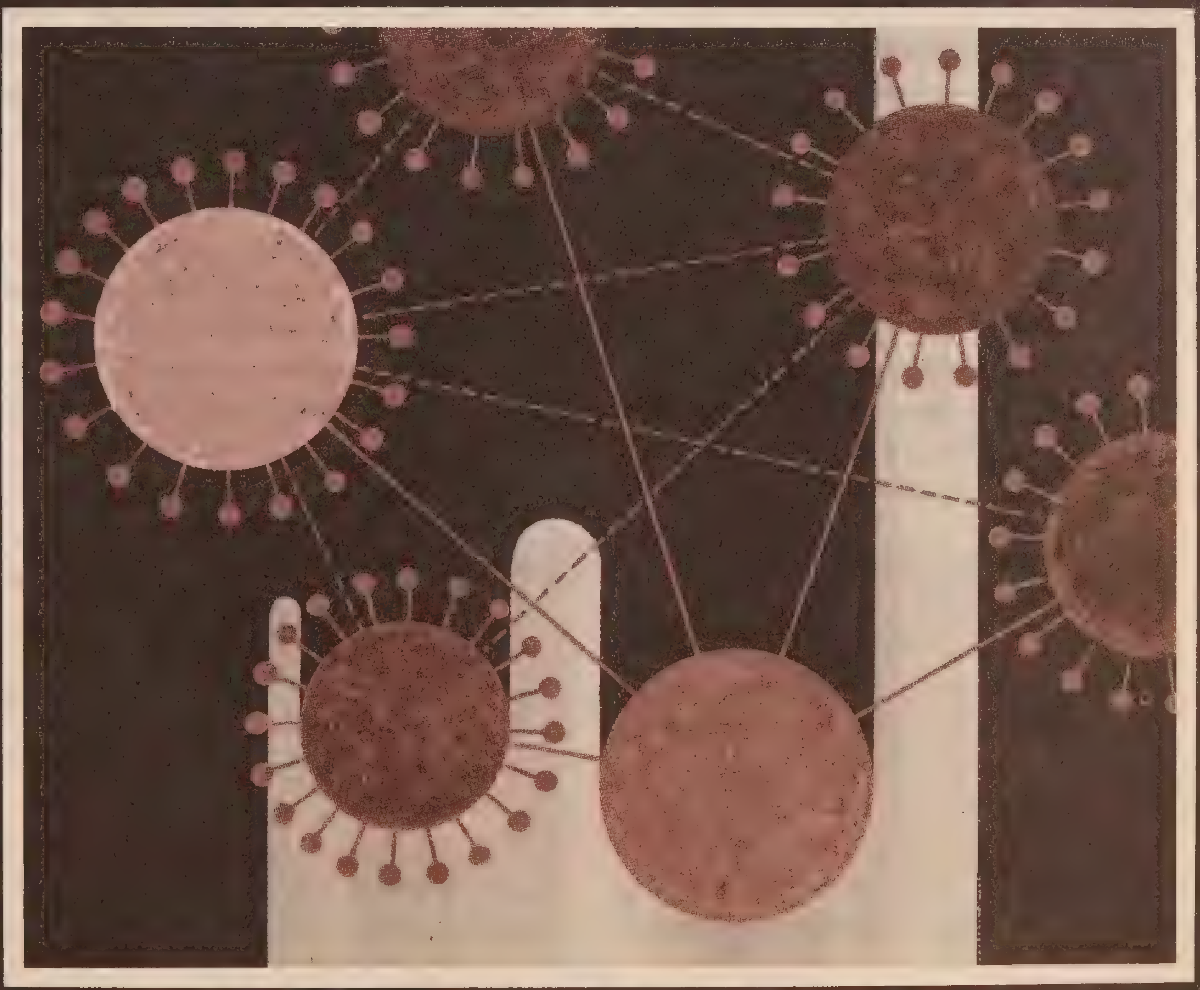
Pitts, J.E. Jr., "Relays and Their Operation Part 1," Radio-Electronics, (February 1951).

GTE LENKURT

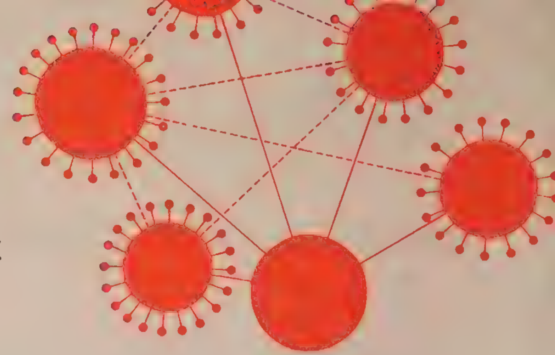
DEMODULATOR

JULY 1974

Some Fundamentals of Telephone Signaling



Even the earliest telephone had to have some way of signaling parties at the other end. Loop signaling is one of the simplest types of telephone signaling, yet it is the stepping stone that leads to more complex forms of signaling arrangements.



Even when Alexander Graham Bell spoke the first complete sentence into the first telephone, "Mr. Watson, come here, I want you," telephone signaling was involved. In this case the sound of Mr. Bell's voice and his words were the signaling which informed Mr. Watson that a message was to be communicated. This method of signaling was adequate for experimental purposes, but in later telephones it became necessary to announce an intent to communicate by means of visual or ringing devices, so that the called party could be alerted at some distance from the telephone instrument.

The basic concept of signaling may seem simple enough, but the process by which signaling is effected may range from the operation of a simple dc circuit, to a complex arrangement of steps performed over a maze of long distance trunk circuits. The whole gamut of signaling is inextricably tied to supervision, dialing, switching, timing, and ringing–subjects which in themselves could cause volumes to be filled with information.

Switching Functions

In a telephone system, all subscriber stations connect to a central office by means of relatively short lines called subscriber loops (see Figure 1). The central office provides the switching arrangements that enable any subscriber station connected to that office to be connected to lines leading to other subscriber stations which are also connected to that office, or to trunk lines that lead to other central offices in the same or distant cities (see Figure 2). Switching equipment at the central office can be either manual or automatic depending on whether switching is performed mainly by humans or by mechanical or electronic devices.

Manual Switching

Manual switching is performed by operators at central office switchboards. All subscriber lines in a local single office exchange terminate in at least two jacks on a switchboard. The jack mounted in the lower part of the switchboard panel (the answering jack) is associated with a signal lamp which lights when a subscriber wishes to attract the attention of the operator. The calling jack, which is mounted in the upper panel of a switchboard, is used by the operator to complete a connection to a subscriber's line when it is called by another station. In other than very small offices, each subscriber line appears at many other jack locations along the switchboard. The terminations are multipled so that any line is within reach of any operator.

The answering jacks and associated signal lamps in a multiple switchboard are divided among the operators, each of whom is responsible for answering signals from an average of two hundred subscribers. An operator is warned against plugging into a line that is busy by a "click" which is heard in the receiver of the headset

when the plug touches the front portion of the calling jack.

Manual switching and automatic switching ultimately perform the same function—that of connecting the calling party to the called party—but the techniques for doing this are different, and the type of equipment used is entirely different.

Dialing Function

The address or dialing function in a telephone system directs the operation of the switching equipment in the automatic offices. Consequently, the evolution of the various switching systems has brought about changes in address signaling techniques. Address signals originate at the telephone dial, and consist of a train of dc pulses corresponding to the number dialed. Modern "touch calling" systems which use keys or pushbuttons instead of a rotary dial, employ tones at different frequencies rather than dc pulses.

In the step-by-step switching systems, the switching equipment responds directly to the dc pulses. However, in panel and crossbar switching systems, the switches cannot be controlled directly by the dial pulses, so these systems require a device known as a "register/sender" which stores the dial pulses and then controls the movement of the switches.

Four basic methods commonly used to transmit address or dialing signals for use by the various switching offices are dial pulsing, revertive pulsing, panel call indicator (PCI) pulsing, and multifrequency (MF) pulsing. Dial pulsing is the most commonly used method of transmitting address information—the numerical value of each digit is represented by the number of pulses in a train (one pulse represents 1, ten pulses represent 0). Dial pulsing is used in all types of switching offices.

Revertive pulsing was originally developed for use in panel switching offices. In this type of pulsing, the address pulses are not transmitted by the originating office. When a call is made, a loop to the distant office is closed. This starts the movement of a panel selecting switch at the distant office. As the selecting wipers pass each terminal, a commutator transmits pulses back to the register/sender at the originating office. When the proper number of these revertive pulses, corresponding to the called number, are received by the register/sender, a signal is sent back to the distant end to stop the movement of the selector. Revertive pulsing is used in certain crossbar offices as well as panel offices.

Panel call indicator (PCI) pulsing is a method of transmitting address signals between a dial office and a manual office. This technique converts pulses received from a dial office to lamp indications which appear on a switchboard. The switchboard operator then connects the incoming call to the called number and rings the subscriber.

Multifrequency (MF) pulsing transmits address pulses between switching offices. Digital information is transmitted in the form of short tone

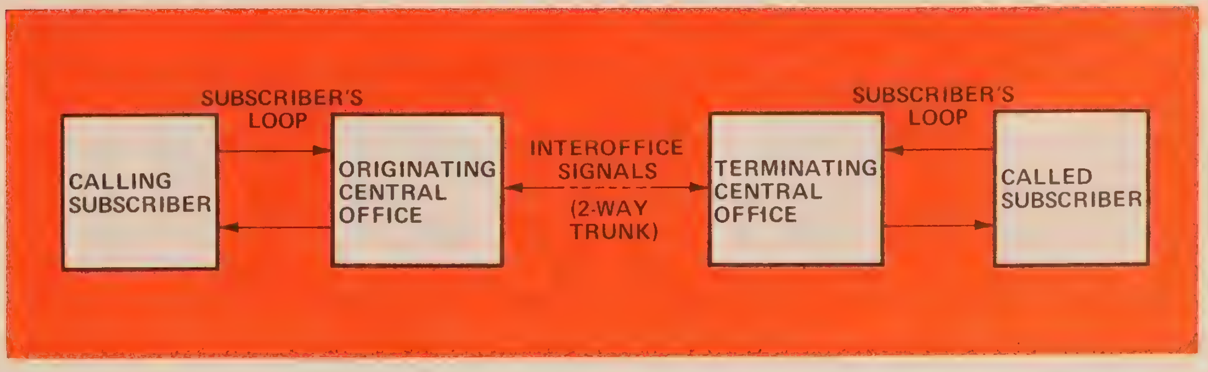


Figure 1. The subscriber's loop connects the telephone to the central office.

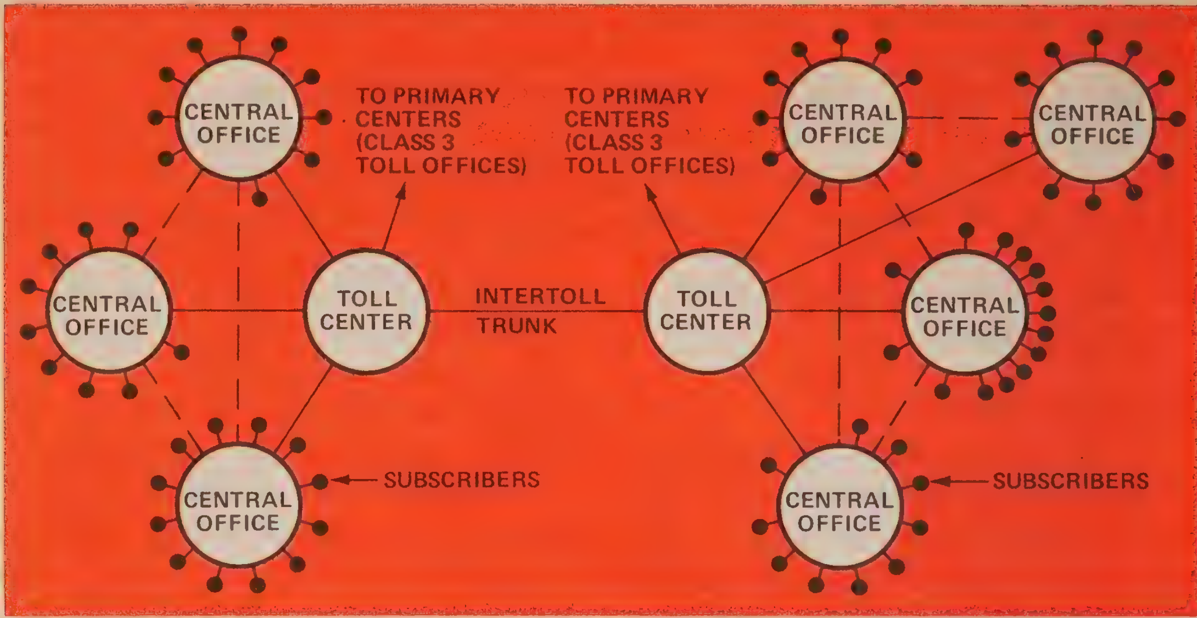


Figure 2. Subscriber stations are connected to other subscribers by a network of trunk lines.

bursts. Six signaling frequencies are used, each digit being represented by a combination of two of the six frequencies. The signaling frequencies fall within the speech band and are processed through the trunk in the same manner as speech signals. A form of multifrequency pulsing can also be used over subscriber loop circuits that use pushbutton telephones instead of the conventional rotary dial.

There are essentially two fundamental techniques used to derive signaling paths on telephone trunk circuits—loop signaling and E&M signaling. Loop signaling requires a dc loop, and is the method used in all subscriber loops and in most short-haul 2-wire trunks. E&M signaling is used with both ac and dc signaling systems, on 2-wire or 4-wire physical trunk circuits and on carrier derived trunk circuits. E&M signaling is standard for use in all intertoll trunks.

Subscriber Loop Signaling

A basic understanding of telephone signaling begins with dc loop signaling. The telephone loop is the closed circuit that is formed by the subscriber's telephone and the cable pair, and any other associated conductors necessary to make the connection to central office switching equipment. Loop signaling takes place over a telephone loop, and requires a metallic dc circuit to operate. Loop signaling systems generally signal by altering the current on the telephone loop. For example, the current flow may be made to change directions, change value, or turn on and off. These changes in current flow are detected by marginal or polar relays at the distant end of the trunk.

Loop signaling is one of the simplest forms of signaling in the telephone network, and is used in certain exchange trunks, short-haul, toll-connecting trunks, and one-way dialing toll trunks where 2-wire voice-frequency circuits are employed. The dc signaling current flows over the same conductors used for voice transmission. Some loop signaling methods include high-low, wet-dry, reverse battery, and grounded battery. Of these, reverse battery and battery-ground signaling (a variation of reverse battery signaling) are most commonly used.

The types of signals that are normally transmitted and received over subscriber's loops can be classified into three categories of local signals: super-

visory signals, information signals, and control signals. Supervisory signals are initiated when a subscriber makes a request for service, retains or releases a connection, or recalls the telephone operator. Control signals provide directions for establishing the desired connection. Information signals notify subscribers as to the progress of their calls, and include audible ringing, dial tone, and line busy signals.

Common Battery Supervision

A type of supervision most often used in a subscriber's loop is "common battery supervision." The development of the common-battery system established the dc source in the central office, and eliminated the need for a hand-cranked magneto and dry cell battery at each telephone station. The central office battery performs essentially two functions; it puts current through the telephone microphone transmitter so that the speaker's voice will change the dc current when the sound waves strike the diaphragm in the microphone, and also gives on-hook or off-hook information to the central office. The telephone receiver does not require dc, since it is an ac device only.

In a common-battery system, dc power is supplied by the central office in the form of 24V or 48V. This "talking battery" supplies power for the telephone transmitter as well as power for signaling and ringing of telephone instruments. When the subscriber lifts the handset from its hook, a dc path is completed through a loop which consists of the customer's telephone, a dc source (battery) in the central office, and a line relay. When the loop is closed, the relay is energized, thus connecting the subscriber's line (or loop) to the operator's switchboard, or to automatic dial equipment. When the calling subscriber is finished with the call, the handset is replaced on the hook, thus breaking the circuit and allowing the relay to release the line.

Reverse Battery Signaling

The tip, ring, and sleeve (sometimes called C-lead) designations that make up a telephone line are derived from names associated with early switchboard systems. The subscriber's telephone loop was connected to trunk circuits by means of plugs that fitted into jacks located on the switchboard panels. The plug connectors bore the designations tip (T), ring (R), and sleeve (S), as shown in Figure 3. The tip and ring leads were used for voice transmission, while the sleeve was used as the control lead and make-busy lead. These designations remain today for lead reference (battery, ground, control), even though a manual switchboard may not be used with a system. To detect whether the called party's instrument is on-hook or off-hook, some means of supervision is required by the central office. Reverse battery signaling is widely used for this pur-

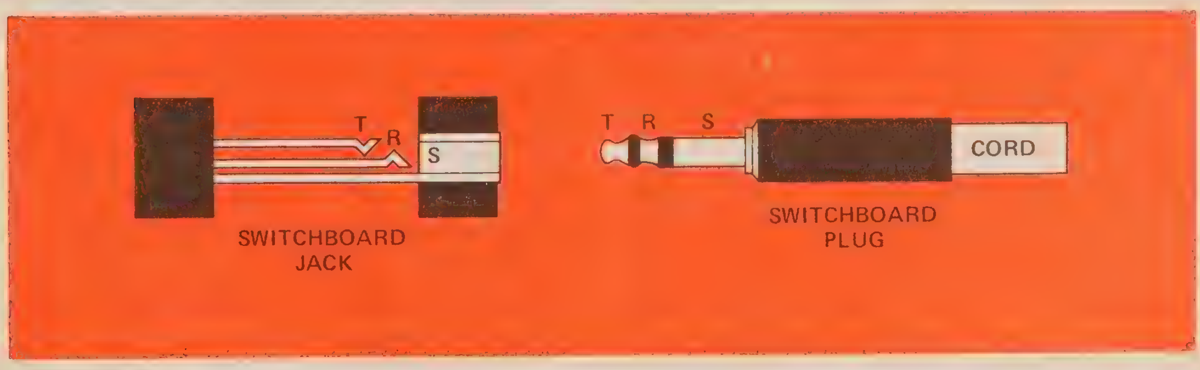


Figure 3. The tip (T), ring (R), and sleeve (S) designations of a telephone line originated from the contact descriptions on a telephone plug.

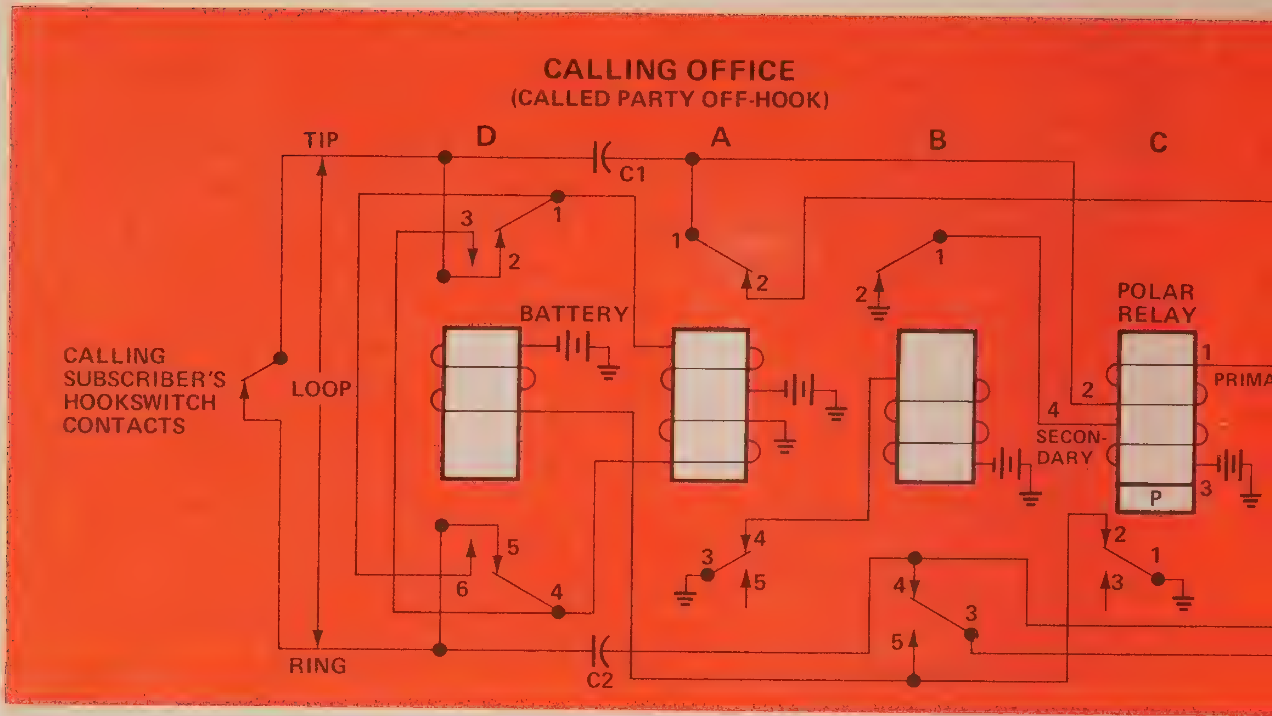


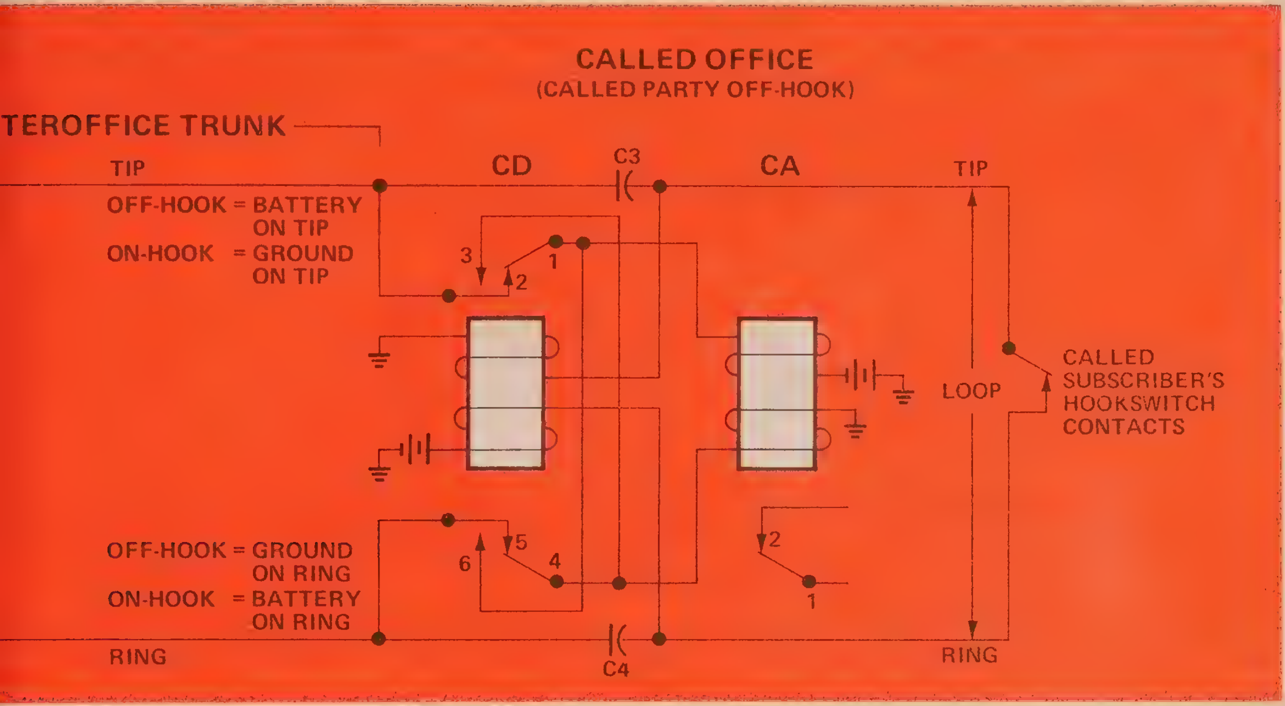
Figure 4. Trunk circuit using reverse-battery signaling.

pose. Reverse battery signaling is so called because battery and ground are reversed on the tip and ring conductors to change the dc signal in the direction of the calling end, from on-hook to off-hook. By means of current reversal, reverse battery signaling provides answering supervision (an indication that the called party has answered the call).

In crossbar and step-by-step switching offices, when an interoffice trunk is idle, there exists battery on the ring and ground on the tip of a telephone line. When the called party answers, talking battery is supplied from the central office. In answering a call, the called party operates a relay as the telephone goes off-hook. The operation of the relay reverses the polarity of the battery back to the originating office. The reverse current notifies the operator (if it is an operator-handled call), by turning off a lamp on the switchboard, that the called party has answered. The reverse battery action also gives notice to message accounting equipment that the called party has answered, and therefore, that the call was completed. This is necessary to enable proper charges to be made to the calling party. When the telephone is returned to an on-hook condition, current polarity again reverses.

In the trunk circuit of Figure 4, all dialing has been completed and the called party has answered. Relays A, B, C, and D are part of the calling party's trunk circuit. Capacitors C1 and C2 couple the voice frequency currents from the subscriber loop to the trunk circuit, but separate the dc control circuits. Relays CA and CD form part of the called party's incoming trunk circuit, while capacitors C3 and C4 couple the voice frequency currents from the trunk circuit to the called party's loop, but separate the dc control circuits. Relays A and B have no direct function in the battery-reversal operation, and remain as shown during the reversal process.

In the circuit of Figure 4, the battery polarity on the calling subscriber's loop is controlled by the called party. When the latter answers, relay CD operates. Contacts 1 through 6 of this relay act as a polarity reversing switch to reverse the trunk battery polarity furnished by relay CA. In the unanswered condition (CD released), the trunk has battery on the



ring and ground on the tip. Terminal 1 of the primary winding of polar relay C, is connected to battery furnished from the called office, due to the released condition of the CD relay. Battery is furnished from the CA relay via contacts 5 and 6 of the CD relay over the ring conductor, through contacts 3 and 4 of the B relay to terminal 1 of the C relay primary. Ground is furnished to terminal 2 of the C relay via the companion path on the tip conductor (contacts 2 and 3 of the CD relay from the ground winding of the CA relay). Under this condition (on-hook), the two windings of the C relay are opposing and the relay is released (contacts 1 and 2 open). The D relay is released and the calling subscriber's loop has ground on the tip and battery on the ring, via the D relay contacts 2, 3 and 5, 6.

When the called party goes off-hook, the telephone closes the loop to operate the CD relay, causing the battery furnished by the CA relay to reverse polarity on the trunk circuit via contacts 1, 2 and 4, 5 resulting in battery on the tip and ground on the ring of the trunk. This condition causes the C relay to operate, since the battery polarity now applied to the 1-2 winding is aiding that applied to the 3-4 winding. When contacts 1 and 2 of the C relay close, the D relay operates, closing contacts 1, 2 and 4, 5, opening 2, 3 and 5, 6. This results in battery polarity reversal to the calling party's loop, exactly as occurred upon operation of the CD relay in the trunks battery polarity. Other relays associated with the loop, detect this reversal for supervision or message-accounting purposes.

Signaling is much more than ringing a bell at the subscriber's telephone. The signaling process is dictated by the functions of supervisory, information and control signals. Signaling begins when a caller picks up his telephone and continues through the dialing process, and extends through those switching functions that ultimately connect the caller to the called party. Signaling functions, such as those which take place in subscriber loop signaling, play an important part in the overall plan of telephone signaling, since they are the starting point of what could be a series of complex steps in the signaling process of a telephone network.

BIBLIOGRAPHY

"A Review of Telephone Signaling," *The Lenkurt Demodulator*, July 1959.

Breen, C. and C.A. Dahlbom. "Signaling Systems for Control of Telephone Switching," *The Bell System Technical Journal*, Vol. 39, (November 1960), 1381-1443.

From Semaphore to Satellite. The International Telecommunication Union, Geneva, 1965.

Notes on Long Distance Dialing, American Telephone and Telegraph Company, 1968.

Principles of Electricity Applied to Telephone and Telegraph Work, American Telephone and Telegraph Company, June 1961.

"Signaling Over Telephone Trunks," *The Lenkurt Demodulator*, July 1966.

Talley, David. *Basic Telephone Switching Systems*. New York: Hayden Book Company, Inc., 1969.

GTE LENKURT

DEMODULATOR

APRIL 1973

COMMON CHANNEL SIGNALING SYSTEMS

Common channel signaling systems offer improved transmission quality, reduced radio loading, and protection against talkdown and loss of billing due to unauthorized foreign tone sources.

In the telecommunications industry, signaling is the process by which a caller on the transmitting end of a telephone system informs a particular party or listener at the receiving end that a message is to be transmitted. But, signaling is also a two-way process which includes supervision. Supervision informs the caller that the called party is ready to talk, that his line is busy, or that he has hung up. Supervision is also that part of the signaling which holds the path together while the conversation goes on. There are a myriad of signaling techniques and a list of them reads like a table of magical incantations, and they are sometimes just as mysterious. Some of these techniques include E and M signaling, reverse battery signaling, duplex signaling, simplex signaling, closed-circuit signaling, open-circuit signaling, dual-tone multifrequency signaling, loop signaling, separate-channel signaling, in-band signaling, and out-of-band signaling. Also, in most telephone systems, signaling is inextricably tied with switching and ringing. Put very simply, signaling can be viewed as the original command and control; switching, as the process which is controlled and which routes the command to its ultimate destination; and ringing, as the audible or visible signal which informs the party at the distant end that intelligence is to be transmitted.

So many different signaling techniques exist in large part because of past necessity to provide adequate signaling for existing line facilities. Also, signaling facilities have been created to comply with the various switching systems manufactured by telecommunications companies. Two of the most widely used signaling techniques are in-band and out-of-band signaling.

Out-of-Band Signaling

Out-of-band signaling uses one or more ac tones which lie within the pass band of the transmission facility, but just outside of the voice band (see Figure 1). Channel filters for out-of-band signaling are designed with an upper cutoff frequency well below the high-frequency end of the voice channel. This reserves a portion of the spectrum for the transmission of signaling tones. Generally, a tone is keyed to convey signaling information in out-of-band signaling systems. Out-of-band signaling is quite economical in that since voice transmission contains very little speech energy at the upper end of the spectrum, and the telephone handset provides part of the effective filtering, filtering requirements may be somewhat relaxed. This makes it possible to provide good quality transmission for relatively low equipment cost, since the greatest cost of carrier systems is in the channel filters. Also, signaling has little chance to interfere with speech when an out-of-band signaling system is used. Because of this, higher signal levels can be used, thereby improving signaling reliability. Some of the problems with

Figure 1. Signaling tones for out-of-band signaling lie outside of the voice band.

out-of-band signaling are that it not only requires a method of coordinating voice and signaling channels, but it also requires a repeater at the end of each link.

In-Band Signaling

With in-band signaling, signaling tones are transmitted within the speech band, usually 1600, 2400, or 2600 Hz, although 2600 Hz is more widely used (see Figure 2). In-band signaling is very flexible in that speech and supervisory signals share the same transmission facility, but at different times. An in-band system is arranged so that supervisory signals are on the line only before and at the termination of a call. And, in-band signaling, in contrast to out of-band signaling, does not require the use of repeaters to reproduce the dc pulse at an interconnect point. One of the objections to in-band signaling is that the signaling tones which lie within the voice band may cause "talkdown" (false signaling due to voice energy). Conversely, the audible supervisory signaling tones cannot be used during conversation as they can with out-of-band signaling.

Guard circuits which distinguish between speech and signaling tones can usually prevent most talkdown problems. When voice frequencies similar to the signaling tone occur, the guard circuit detects that the signaling tone frequency is caused by speech, not signaling, and therefore prevents signal circuit response. In-band signaling is more complex and expensive than out-of-band signaling, and also requires a more elaborate filtering system.

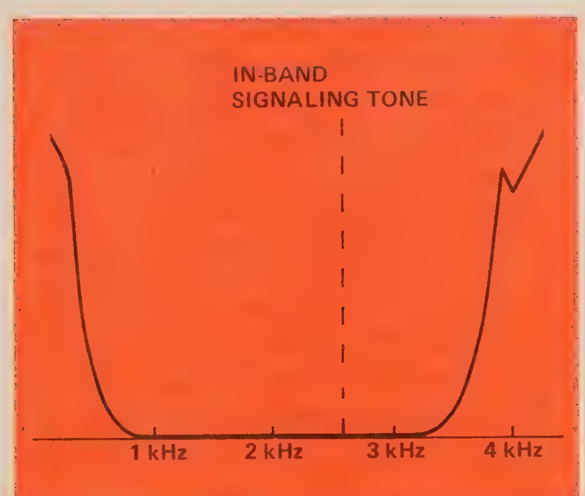
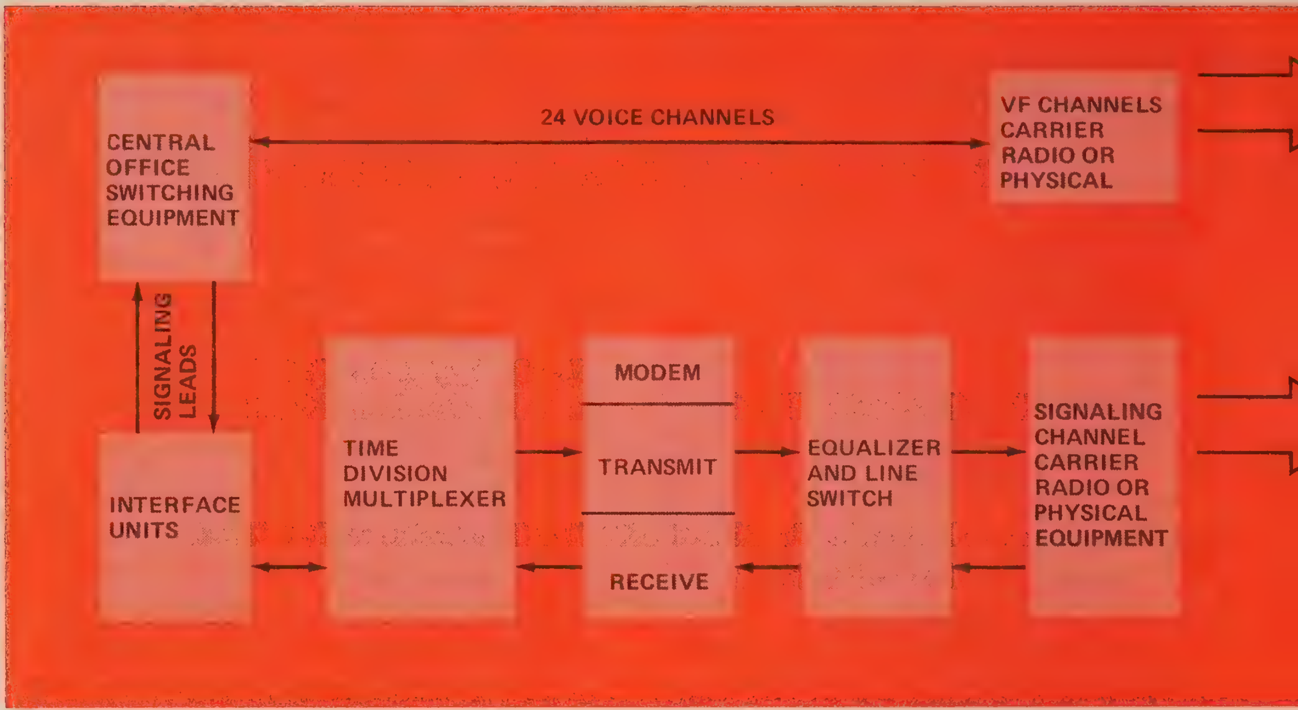


Figure 2. In-band signaling employs tones within the voice band.

Common Channel Signaling Systems (CCSS)

Although in-band and out-of-band signaling systems are efficient methods of signaling, both require extensive use of filters and associated circuitry to prevent signaling tones from interfering with the voice band and other equipment such as SATT (Strowger Automatic Toll Ticketing). In the interest of economy, interoffice common channel signaling systems such as GTE Lenkurt's 11A CCSS have been recently introduced. CCSS is an alternative to in-band, single-frequency signaling. It is not a new concept – the Lenkurt 81A exchange trunk carrier, introduced in 1958, used a similar signaling technique. Before that, the Lenkurt 44B signaling system, introduced in the early 1950's, used a common-channel principle in its multitone signaling. The basic concept of common channel signaling is that all signaling



for a number of voice paths is carried over one common channel, instead of within each individual channel.

The greatest advantage offered by CCSS is a savings to both telephone company and customer. CCSS eliminates talk-down and talk-off problems. It eliminates the need for signaling augmentation (greater gain for signaling tones during dialing). It improves the performance of the voice channel because noise generated by in-band signaling components is eliminated. CCSS is also economical in that special filtering is not necessary as in other signaling schemes. And, of more recent interest, CCSS *can* eliminate the "phone phreak" problem which now plagues many telephone companies.

One GTE Lenkurt CCSS terminal can provide signaling for 24 voice channels. It can accommodate physical line, carrier, or microwave equipment. CCSS is a means of interoffice communication, with equipment located in the central office. Figure 3 shows a block diagram of a typical common channel signaling system. The voice paths come out of the central office switch termination and go directly to the carrier channel banks. The signaling for these voice paths enters the CCSS interface units by loop signals or E and M leads. The M lead transMits "on hook" or "off hook" signals to the CCSS equipment, and the E lead recEives on-hook or off-hook signals from the CCSS equipment. Each CCSS terminal contains four gate units, each of which is capable of handling six signaling channels.

The function of the interface units is to accept various signaling schemes and make them compatible with the CCSS equipment. Signaling between telephone systems is done on an on-hook or off-hook basis, which is a one or zero binary situation, and as such, is well suited to the sampling used in time division multiplex techniques.

A signal gate unit in the time division multiplexer connects to the signaling portions of the interface units, and samples the transmit inputs for on-hook or off-hook conditions at approximately 2400 times a second. Each sample is sequentially fed to a 2.4 kilobit modem. In this way, 24 parallel signals are serialized into one TDM (time division multiplex) stream. At the modem, the TDM stream is transformed into an FSK (frequency

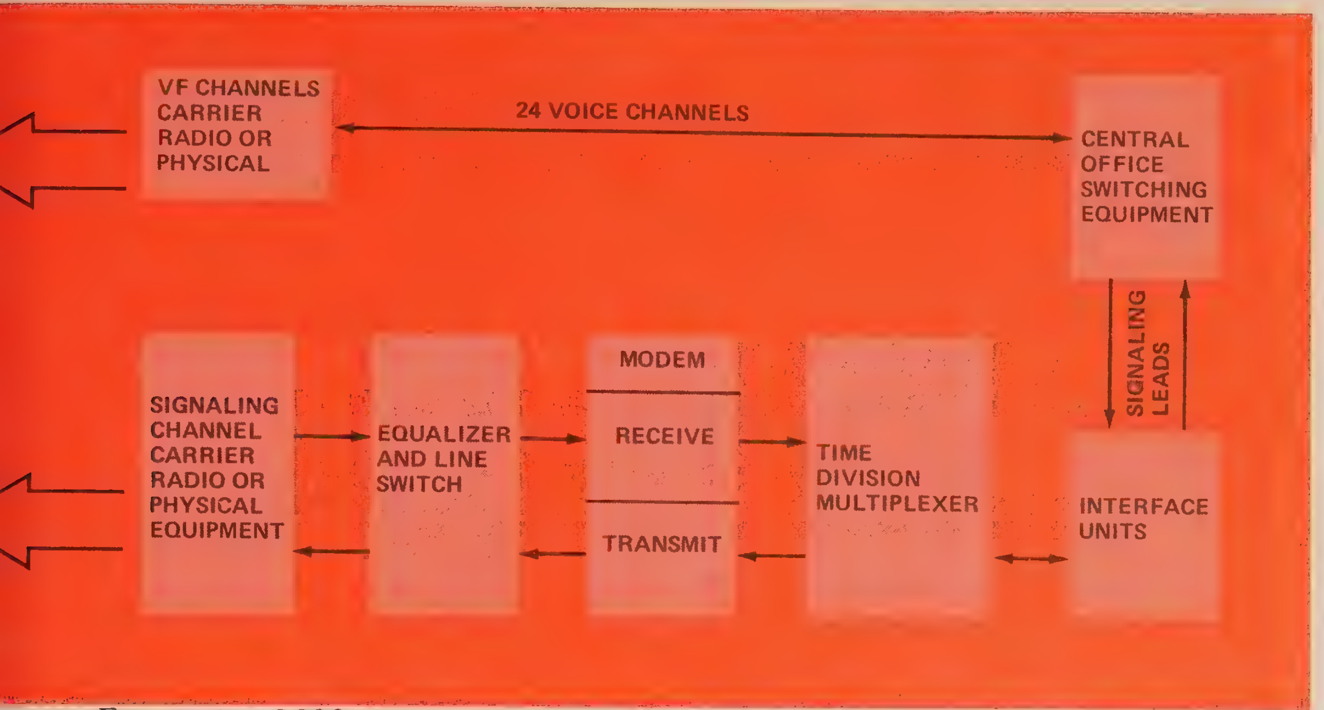


Figure 3. CCSS can accommodate signaling for carrier, radio, or physical equipment.

shift keying) signal within the voice band, which in turn can be put over any standard voice channel, for the purpose of signaling. The reverse of this procedure takes place at the other end of the line.

The Alternate Channel

One means of protecting the CCSS signaling data stream is provided in the form of an alternate channel. Occasionally, a channel will become unsuitable for transmission because of impulse noise. While impulse noise does not affect voice conversation to any large extent, it may introduce a high enough error rate in the signaling channel to interfere with the signaling function. To guard against this, two "quiet" channels (preferably in different groups) are selected when the CCSS is initially installed. Under normal conditions, one of these will be used to carry the signaling and the other will operate as a normal, traffic-carrying voice channel. An error detector monitors the error rate in the signaling channel. This circuit takes into consideration the noise characteristics of channels. That is, that errors due to noise occur in bursts — they are not always present. The bursts are randomly spaced and random in length. Approximately 95% or more of these bursts are less than 50 milliseconds long, and the time between them is approximately five seconds or less. The error detector ignores the first 50 milliseconds of any kind of error occurrence, then begins to count errors for a period of ten seconds. If less than three errors occur during this period, the system goes back to normal operation. If more than three errors occur, the line-switch unit automatically switches the signaling-channel information to the alternate voice channel, and the alternate-channel information to the signal channel. This switch takes approximately 10 milliseconds, which is fast enough that a customer on the alternate channel will not know that a switch has occurred. And, since impulse noise does

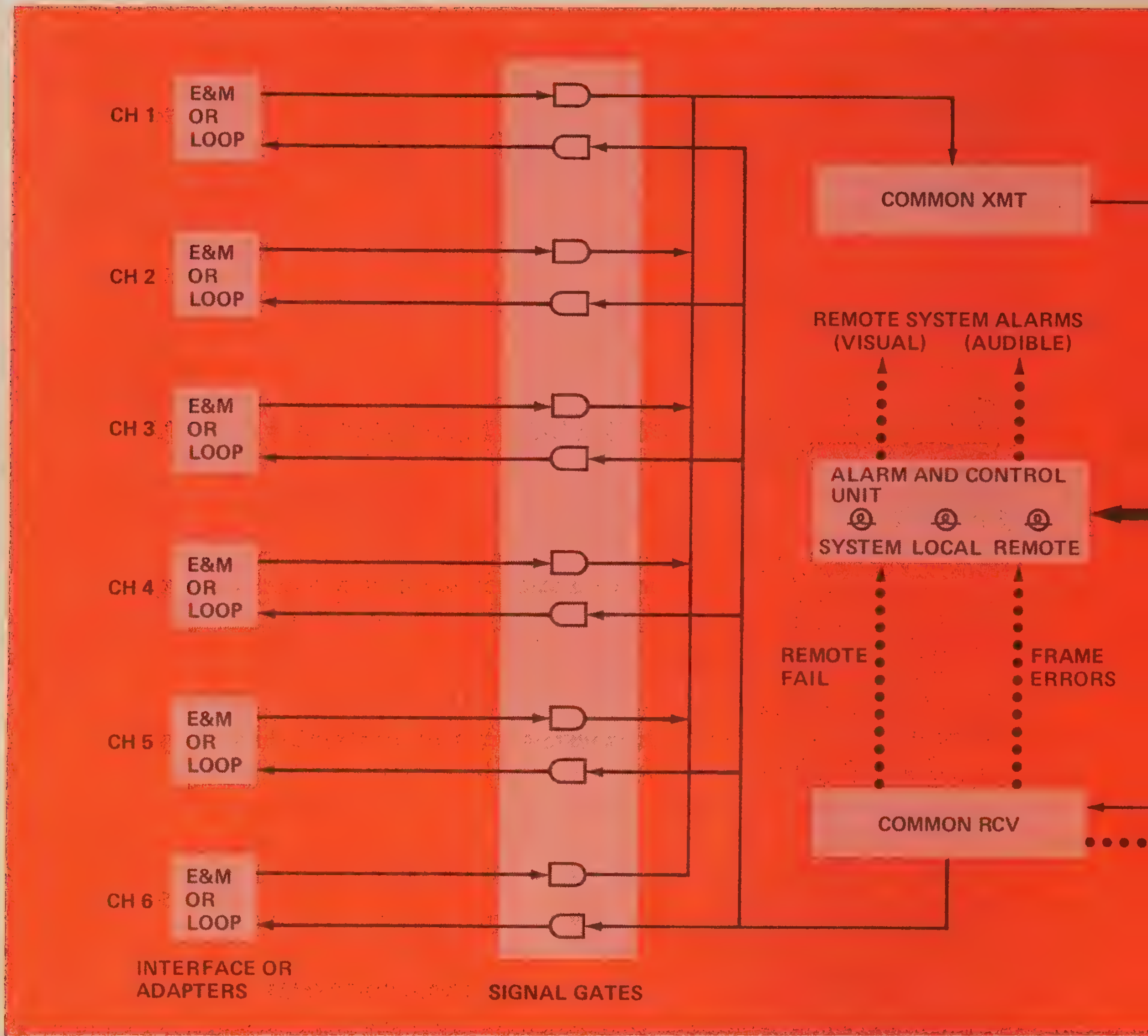
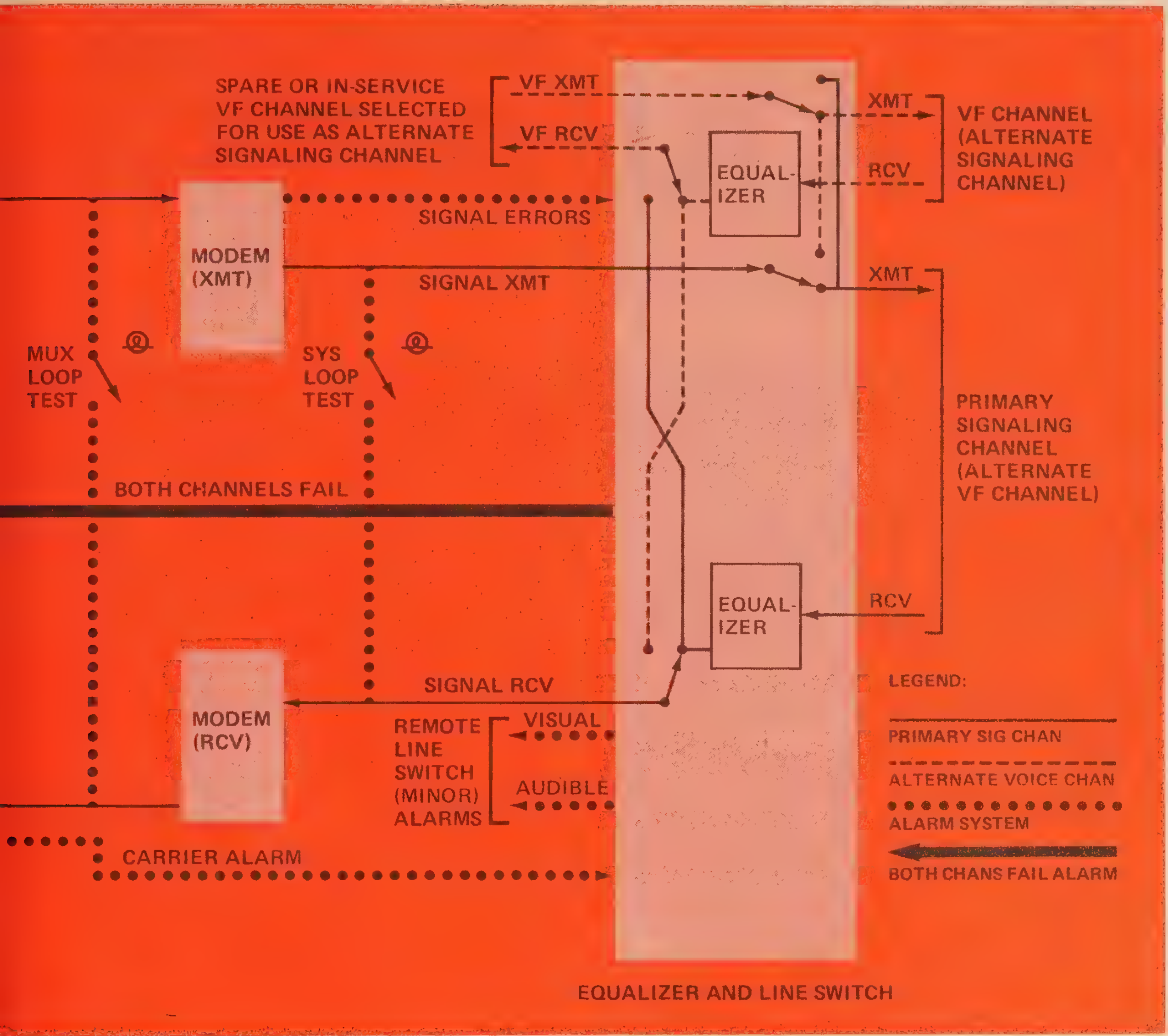


Figure 4. The CCSS alternate-channel protection function is initiated when an increase in error rate or a channel failure takes place.

not seriously impair voice, the customer can continue his conversation. This technique takes advantage of another peculiarity of impulse noise, and that is that noise does not necessarily occur in all channels at the same time. So, in effect, it is quite likely that when there is noise on one channel, the alternate channel will not be undergoing a similar effect. A "minor" alarm is sounded to indicate that a switch has occurred.

In the event that both channels are bad, and this would be a rare situation, a "major" alarm system takes over — bells are rung, lights are turned on, the system is made busy, and a technician is sent to remedy the situation. Figure 4 depicts the CCSS alternate-channel protection system. The thin black lines trace the primary signaling channel, the dotted lines show the alarm paths, and the dashed lines show the alternate channel. When a switch is effected, the signaling data which was initially traveling over the primary signaling channel, now is put on the alternate channel. At the same time, the customer, who was on the alternate channel, is placed on the signaling channel. The alarm path is activated if both channels fail (a major alarm), or when a line switch takes place (a minor alarm). Another type of protection for the CCSS is in the form of continual monitoring of one terminal



by the other, with an alarm being given in case of malfunction.

Phone-Phreak Phree

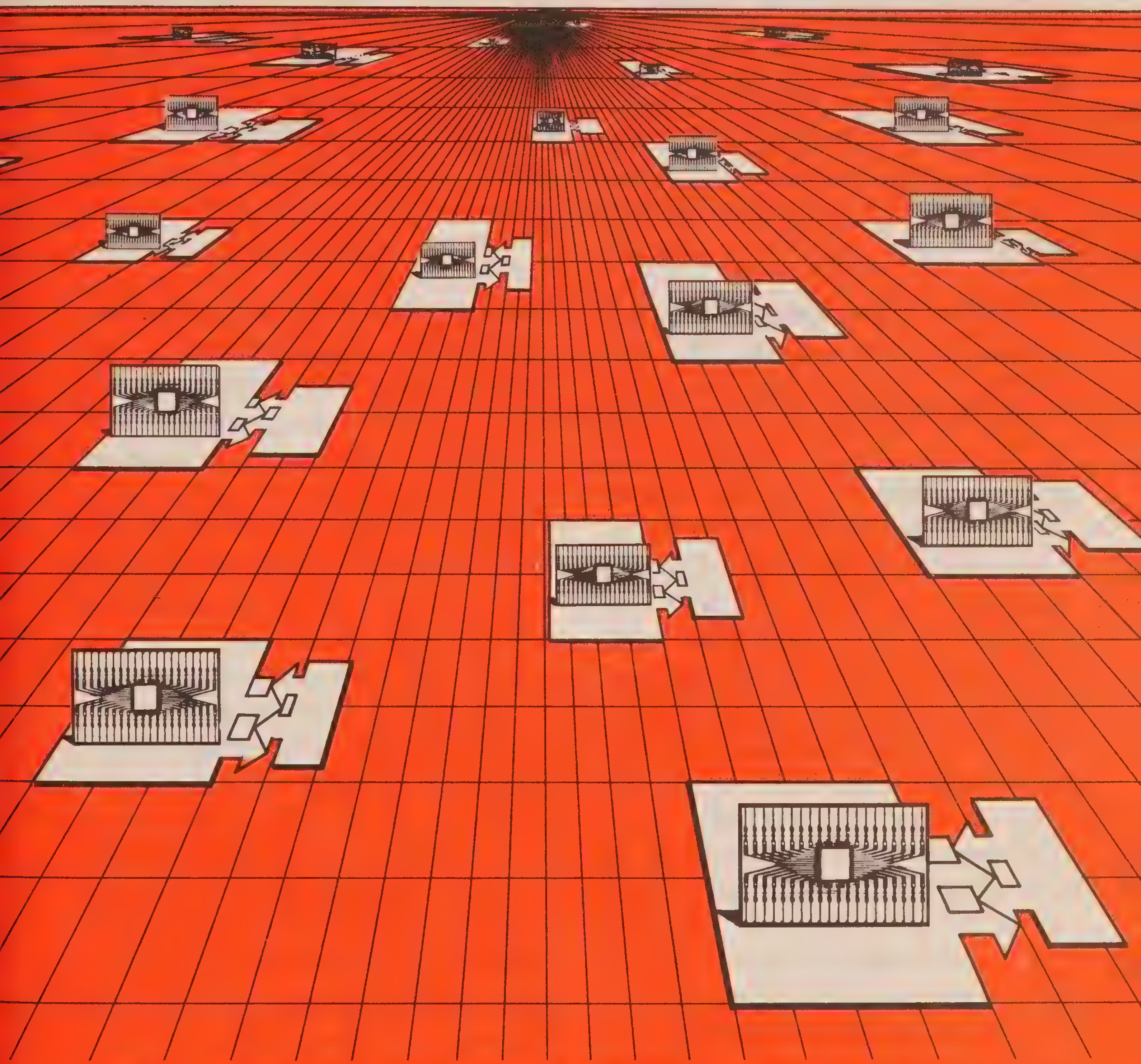
Telephone abuse by the so called "phone phreak" is one of the recent problems that has been frustrating telephone companies, and is costing an untold amount in lost revenue. The phone phreak is an individual who has enough knowledge about telephone system operation to access the system by generation of generally used signaling tones. By generating tones of the correct frequency and at the correct time, the phone phreak can make long distance calls without being charged. This kind of telephone abuse is most commonly found in systems that employ in-band or multifrequency signaling. Since the CCSS technique completely separates the signaling from the voice channel, false tones cannot be used to access the telephone system. The only stipulation here is that CCSS be used *end to end*; a tandem link with a tone-signaling system would make the system vulnerable to the phone phreak.

Common channel signaling is an improvement over most conventional signaling — both in economy and quality of transmission. It is a concept which will become increasingly prevalent in future systems. And, with systems such as the 24-channel CCSS, small offices as well as large can install them and add more as the need arises.

GTE LENKURT

DEMODULATOR

NOVEMBER 1973



THIRD GENERATION SUPERVISORY CONTROL SYSTEMS

Like other forms of technology, status monitoring is in a constant state of evolution, which can be seen in the implementation of new solid state technologies, including LSI, and the use of computers as an integral part of supervisory control systems.

The term "supervisory control" has come to mean the process of receiving digital or analog indications from a remote location, and sending a digital or analog command to a remote location. Some of the techniques used to achieve supervisory control are remote indication, telemetry, and telecontrol. Early systems usually included only the remote indication and telecontrol functions, representing only the digital forms of supervisory control, because transmission of binary quantities (on or off) was much easier than the transmission of analog quantities, and the need for analog measurement and control could not justify the high cost of early analog-to-digital converters. Later systems often included additional functions, especially in industrial applications where "process control" (the automatic control of a complex industrial process) was required. Data logging was also used where a historical record of system operation was required.

In hardware technology, supervisory control systems have evolved through two earlier generations, and are entering a third generation. The term "third generation" as applied to a supervisory control system is arbitrarily used to describe a family of systems utilizing LSI (Large Scale Integration) in the component circuitry, as well as having the optional facility to report into and receive controls from a minicomputer, or having a computer system and associated software as part of the system. Such systems, such as the GTE Lenkurt 51 series, are now being employed by users of telecommunications equipment for reporting and controlling of unattended microwave repeaters and for CDO (Community Dial Office) alarm reporting. Oil and gas companies use such systems for transmitting production data and for pipeline monitoring and control. Power utilities remotely control and monitor generating stations and substations from a centralized location, while simultaneously collecting grid data for load and frequency control.

First generation supervisory and control systems utilized electromechanical relays and rotary switches

plus vacuum tube tone transmitters and receivers. This equipment was rugged and bulky but performed well, and there are hundreds of such systems still in service.

Second generation equipment utilizes integrated circuits and discrete components in place of relays, with solid state FSK (frequency shift keying) or AM (amplitude modulation) tone transmitters and receivers in place of vacuum tube types. The results have been a tremendous reduction in size (approximately 5:1) as well as an improvement in reliability with lower first cost and with less maintenance. Associated with the second generation of supervisory and control equipment are data loggers and custom designed minicomputers to gather information from many locations, and present this at a centralized alarm and control center.

Third Generation Families and Concepts

It is logical that a new generation of equipment should incorporate the best features of the previous equipment, as well as use the operating experience gained with these systems to add new and desirable features. The first step is to take a broad overview of individual alarm systems, control systems, minicomputer systems, etc., and to design a true "family" of systems with common modules such as encoders, decoders, and power supplies. The second step is to design the new families for the broadest areas of application by all users, whether in the telecommunications, oil and gas, transportation, utility, or any other field. This is achieved at low cost with the use of LSI (Large Scale Integration) chips, where up to 4,000 resistors, diodes, transistors, and other components can be deposited on a 1/4-inch square silicon wafer. Additional features can be built-in initially at minimum cost.

The third step in the design of new supervisory and control systems is to design them with the option to work into a centralized computer-based reporting and control facility, whether or not one is initially installed. Experience has shown that the high cost of software programming for the custom designed second generation minicomputer supervisory systems could be sharply reduced if a standard series of programs could be written. But, due to the wide diversity of applications for supervisory and control systems, this is not completely possible. What is possible, however, is to provide a broad selection of standard programs for specific functions such as alarm reporting, control with "check-before-operate," plus a "System Builder" program to allow the user's operation and maintenance department to readily add and delete alarm indicators, and control functions, as desired, without requiring more than a minimum knowledge of programming.

Remote Indication Systems

Remote indication systems differ principally in the mode of reporting used. The more common types are continuous reporting, time-slot reporting (polled reporting over a dedicated facility), and dialed reporting over a switched network. The latter is a variation of the polled type of operation.

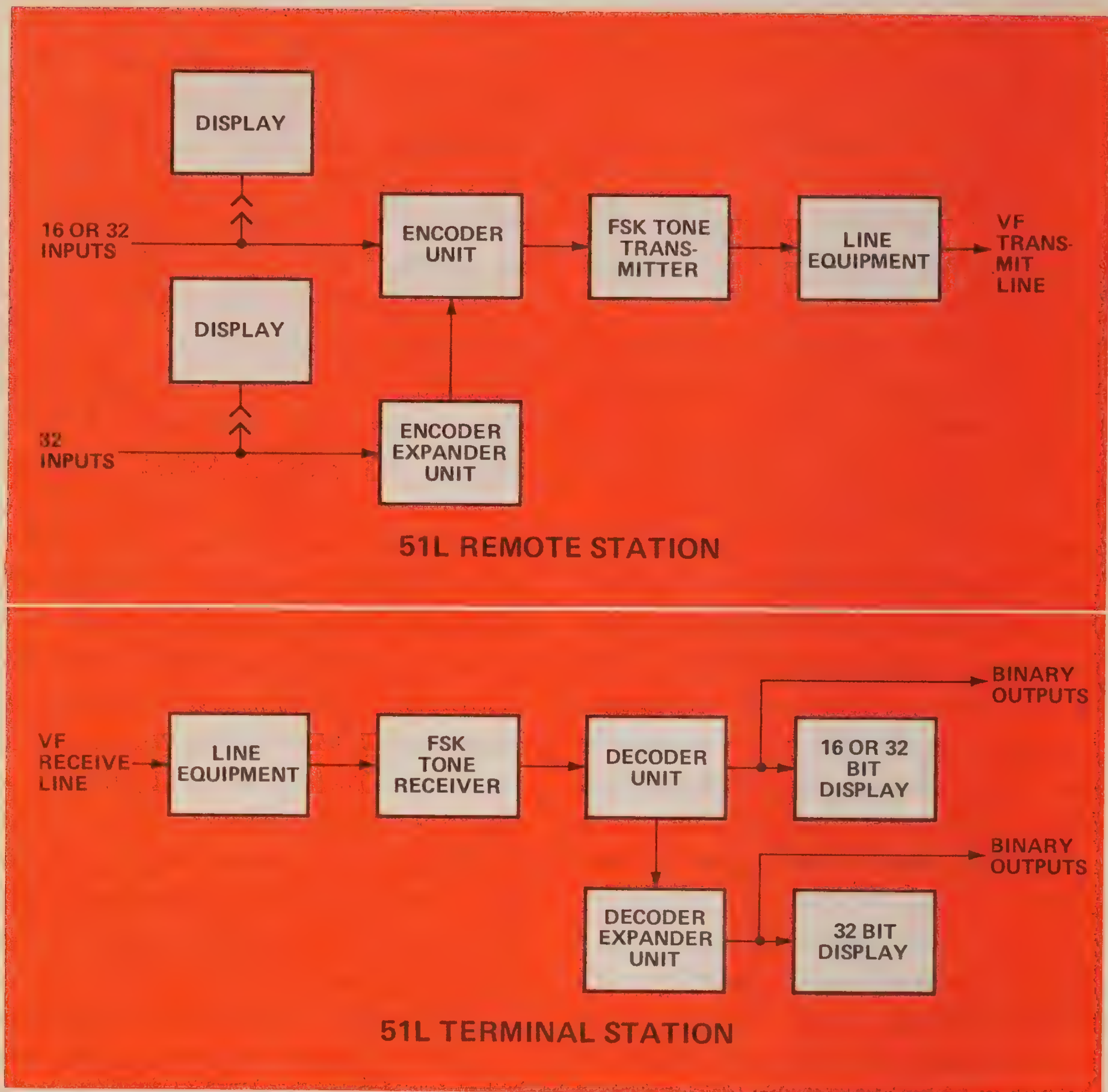


Figure 1. A continuously-reporting system gives an immediate alarm if the tone to the terminal station is interrupted.

Continuous Reporting

In a continuously-reporting system, each remote station reports the status of its input points continuously and automatically over a dedicated facility, using an individual tone channel between each remote station and its associated master station. The term "fail-safe" is often applied to such systems, because absence of tone causes an immediate alarm. Reporting is rapid, since the remote station can report continuously at a high baud ratio, and does not time share with other remote stations. One-way transmission is all that is required, adding more reliability to the reporting of information, as well as allowing a remote station to report to as many master stations as necessary. Modern continuous-reporting systems, such as the GTE Lenkurt 51L (see Figure 1),

provide a wide range of reporting rates and station capacities, and find their greatest application where a relatively large number of status points per station must be transmitted in the minimum time.

Time-Slot Reporting

Time-slot reporting systems use one tone channel to report the status points of a number of remote stations. Since only one tone channel is used, the remote stations must time share the facility; that is, each station is allotted a time slot in which to report its information, hence the stations report sequentially by keying on their carrier and sending the station data, then turning off the carrier to allow the next station to report. Although speed of reporting is much lower than with continuous reporting systems, bandwidth requirements are correspondingly much lower. Consequently, this type of operation is suitable for systems having many remote stations of fairly low number of points each, where system reaction times allow a slower reporting rate. There is also an economic advantage in that a simple master station can receive indications from many remotes, rather than requiring a separate master for each remote as in the continuous reporting type.

To ensure time sharing, the remote stations must be synchronized with each other. This is accomplished in systems like the GTE Lenkurt 51M alarm reporting system by sending a momentary signal from the master to all remote stations after each complete system scan to reset the crystal-controlled counters in each remote (see Figure 2). The important advantage of time-slot reporting as used over a polled system is that reporting ability is not lost if the outgoing synchronizing signal is lost, since the crystal-controlled remote stations will remain synchronized for several hours and will continue to report.

Because each station in a time-slot reporting system must report within a certain time window, station-point capacity is limited. A system such as the GTE Lenkurt 51M may be arranged to report 26 alarms per remote station, allowing up to 63 stations per tone channel, for a system capacity of 1638 supervised points. If more points are required, two or more remote units may be located at the same station, each having its own time slot. Each remote station in the time-slot system is identified by a unique 6-bit station address.

To overcome the main disadvantage of time-slot reporting, a change of state reporting sequence may be used in which any station having a change of state, which might constitute an emergency condition, will report immediately in the time slot being transmitted at the time by sharing the time slot with the designated station. This is possible because each station reports in the center one third of its time window, leaving the other two thirds empty of any transmitted signal. It is the last one third that is used by any station having a change of state.

Polled Systems

Polling refers to the method of operation where the master station

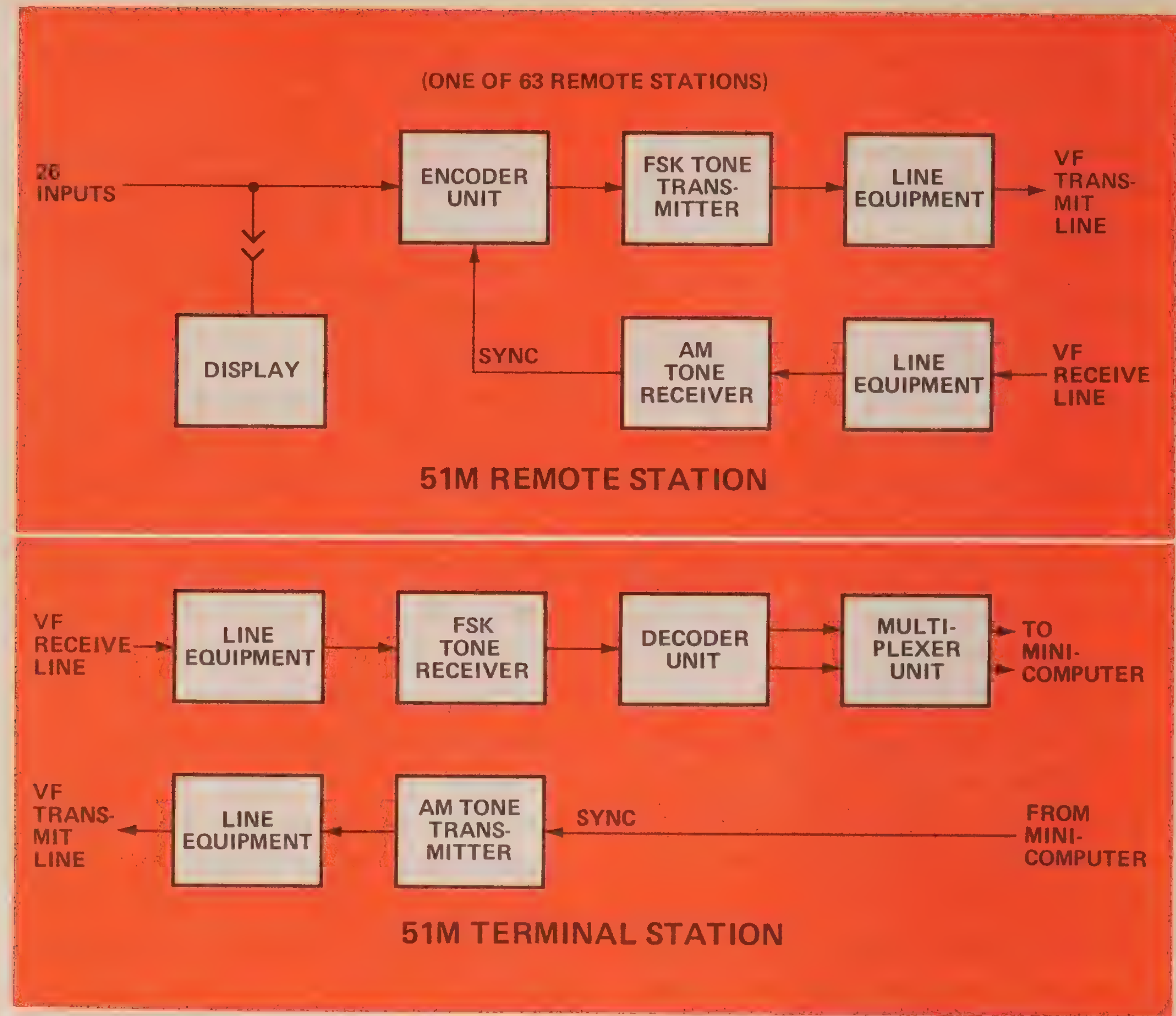


Figure 2. Time-slot reporting systems use one tone channel to report the status points of several remote stations.

sends a request to each remote station in turn to report its status. A two-way transmission facility is required, an outgoing path for the interrogation signal and an incoming path for the station reports. This type of operation has the disadvantage of being disabled if either direction of transmission is lost, and is therefore often considered unsuitable for certain applications. For example, a polled system monitoring a microwave radio system can be disabled by a communications failure just when reports of the failures are vitally needed by maintenance personnel.

Dialed Systems

Dialing remote stations through the telephone switched network is a variation on the polling method of operation. However, instead of requiring a dedicated two-way channel, a non-dedicated telephone link is used which is very economical in terms of channel requirements because the channel is accessed only when needed. The natural application of dialed systems is in the monitoring of community dial offices by the telephone companies.

Unattended Community Dial Offices (CDO's) normally report alarms

via physical facilities to their nearest attended office. To provide 24-hour coverage, simple lamp displays often appear on telephone operators' switchboards. To provide a high service level with the increased cost of "call outs" (when a repairman has to be sent out), it is becoming apparent that much more information is required about the specific conditions in the CDO, and that this information should be analyzed by knowledgeable maintenance personnel so that the correct type of repairman can be sent to perform the job. Centralization of CDO alarm information, particularly in the off hours, is therefore becoming a priority item.

As each CDO is readily accessed via the telephone switched network, a telephone call to a specific number in each exchange can access a continuously running encoder to which up to 32 office alarms may be connected. Each station transmits its address and alarm status to a common decoder at the central location.

Automatic dialing of each CDO may be arranged on a pre-determined period, either utilizing a bank of repertory dialers, storing up to 25 numbers each, or automatic dialing under minicomputer control with numbers stored and output sequentially.

Telemetry Systems

Telemetry, like remote indication, involves the transmission of information from a remote location to a master station for use by the operating personnel. The difference is that where remote indication refers to two-state indications, i.e. on/off or open/close, telemetry refers to analog quantities which can take on a continuous range of values. Earlier systems sometimes transmitted the quantity by analog methods, such as variable frequency, pulse amplitude, or pulse position modulation. Many sources of inaccuracies as well as susceptability to noise made such methods undesirable.

Telemetry systems now transmit analog quantities by converting the quantity to digital form at the originating site, and transmitting the resulting digital signal by methods previously described for remote indications. Using this method, errors can be eliminated through digital code checking so that inaccuracies do not creep in as in analog transmission.

With today's low cost of analog-to-digital converters, telemetry of numerous quantities previously not considered necessary becomes practical. A change in philosophy is taking place in which remote measurement takes a much more prominent place in systems operation and maintenance. Telemetry has long been used for measurements such as pipeline pressures, flow rates, task levels, electrical distribution quantities such as current and voltage, and other operating system parameters.

It is now being used for many other purposes such as preventative maintenance, traffic measurements, environmental data gathering, and other tasks which may not be classified as essential, but do increase the efficiency of the operation and provide services not available previously.

Modern telemetry systems are all solid state and can be arranged to

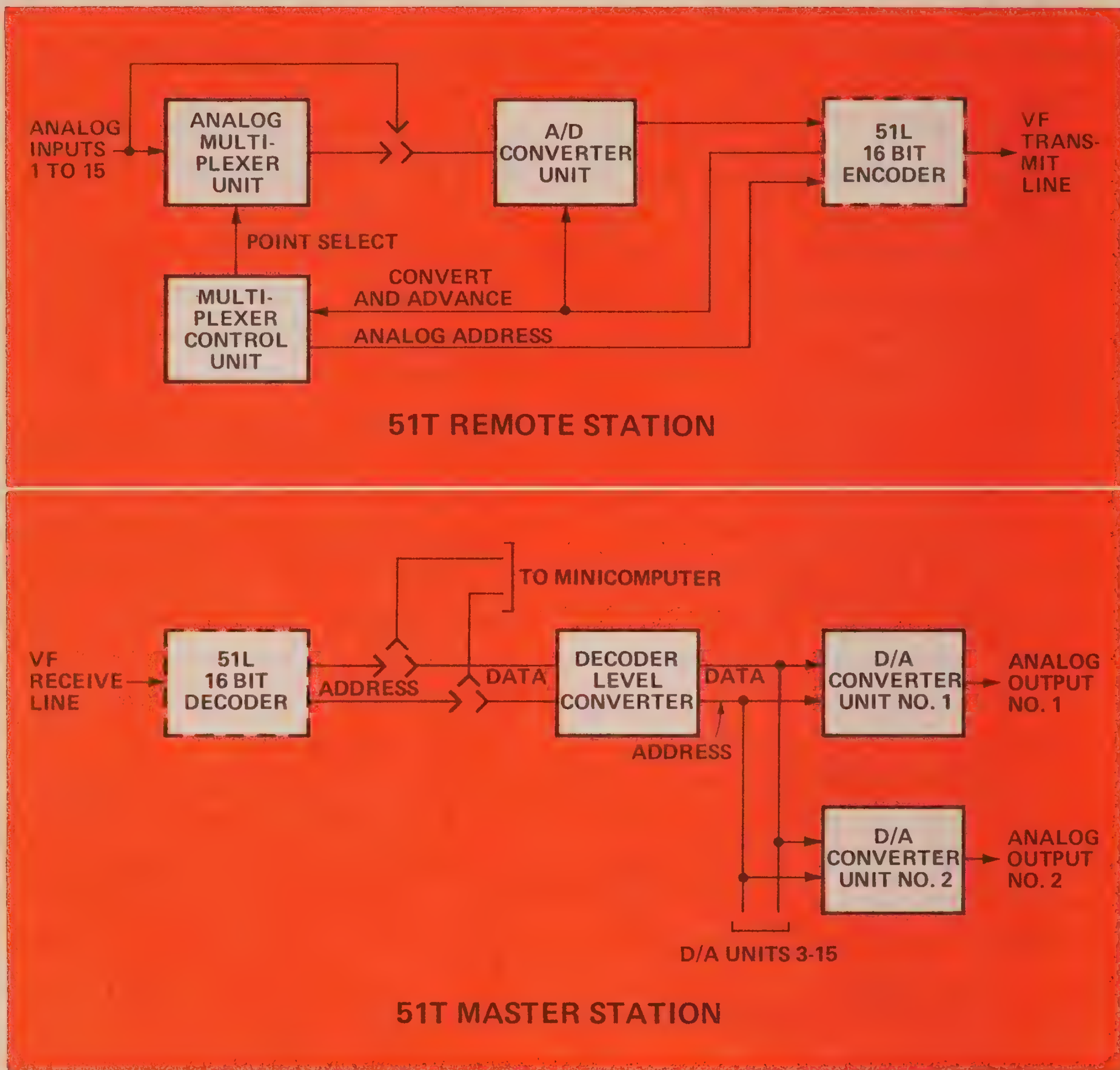


Figure 3. Telemetry systems convert analog quantities to digital at the remote station for transmission to the master station. At the master station the quantities are decoded and displayed in the original analog form.

transmit a large number of different analog quantities very efficiently and at speeds adequate for the process or operation being monitored. A telemetry system, such as the GTE Lenkurt 51T, is an assembly of analog-to-digital (A/D) converters, analog multiplexers, digital-to-analog (D/A) converters, and ancilliary equipment, all designed to work into a digital transmission system such as the GTE Lenkurt 51L remote indication system. In certain applications, where rapid updating of telemetry information is not required, a time-slot reporting system, such as the GTE Lenkurt 51M remote indication system may be used as the digital transmission system. The telemetry system is designed to accept analog data and convert it to digital form for transmission. At the master terminal, the incoming digital informa-

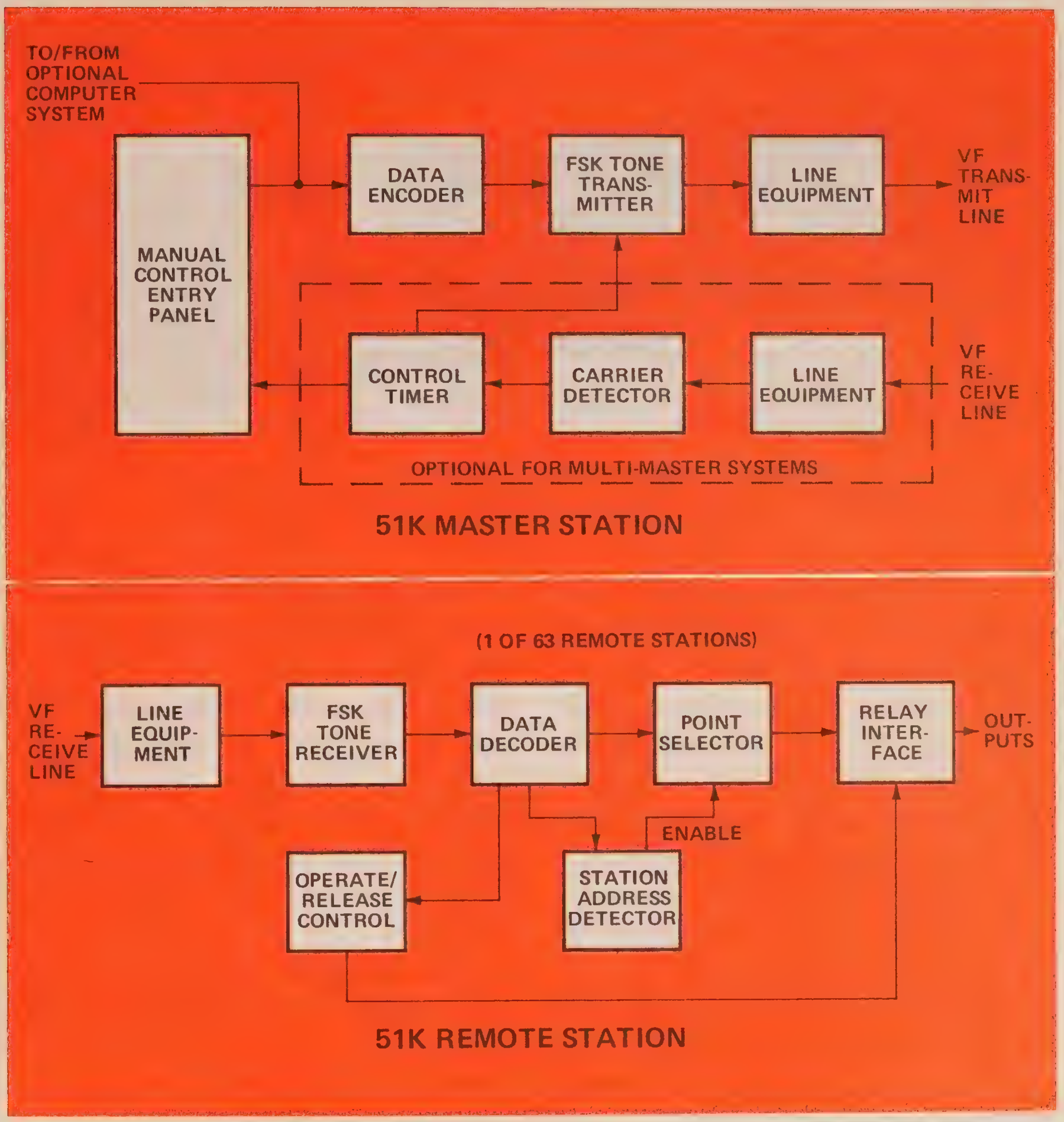


Figure 4. Telecontrol systems transmit control commands to remote stations to perform such functions as activating relays and operating valves.

tion can either be displayed in decimal form, reconverted to analog form, or input into a minicomputer (see Figure 3). The extreme variability of telemetering requirements make great flexibility, mandatory.

Typically, a number of analogs of different characteristics must be converted to one standardized level in appropriate interface units. The conditioned analogs are sequentially scanned by an analog multiplexer with the resulting output being presented to an analog-to-digital (A/D) converter. The digital output, which usually has an 8-bit to 12-bit binary resolution, is transmitted by status systems such as the 51L.

Figure 5. Typical control panel for a telecontrol system.

At the 51L master terminal the output may be processed by a computer and/or displayed digitally. If more than one analog value is being scanned, an analog address is included in the data word. This address is recognized by a decoder which routes the data word into the appropriate readout.

Similarly the data may be routed to digital memory units incorporated in digital-to-analog (D/A) converters. The analog address in this case accesses the appropriate digital memory, which, in turn drives the associated D/A unit, resulting in a continuous analog output from each D/A converter equipped.

Telecontrol Systems

Telecontrol systems are used to transmit control commands to remote stations to activate relays, operate valves and all forms of remote control functions. The term "telecontrol" itself is normally associated with transmission of two-state binary control commands while set-point control is more identified with transmission of an analog control command to a remote station. Both types require a very secure control code to prevent errors that could result in operating the wrong valve or turning off a microwave transmitter. A typical telecontrol system diagram is shown in Figure 4. This system provides control of up to 128 latching or non-latching relays at each of 63 stations. The commands are transmitted to the remote stations via a common FSK tone channel.

The structure of control command allows the remote stations to perform synchronization, timing and noise tests. In addition, double scan protection virtually eliminates false control operation due to transmission errors. The control decoding stations are arranged to inhibit simultaneous operation of more than one control relay. Any attempted simultaneous action may be reported to the control station as "Command Inhibit" indication.

Use of optional latching relays to interface the control decoders to the controlled equipment ensures the mag-

netic storage of all control functions in the event of remote station power failures. These are particularly used with set-point storage.

When required, additional system security is obtained by operating the 51K telecontrol system in conjunction with a status reporting system to provide a "Confirm before Operate." The status lamp on the status reporting system confirms that the correct control point has been selected. The "Execute" button may then be pressed to carry out the control function. A typical control panel for such a system is shown in Figure 5.

Modern supervisory control systems make use of LSI technology to provide "third generation" equipment at lower cost. These systems, through LSI, are able to provide the flexibility and features demanded in the variable and often complex applications seen in the world of supervisory control.

BIBLIOGRAPHY

Dahlbon, C.A., "New Supervisory Control System," *Bell Laboratories Record*, August 1955.

"Remote Control and Supervision of Industrial Operations," *The Lenkurt Demodulator*. September 1955.

Rochat, R.A. and W.O. Hawk, Jr., "Modern Supervisory Control Systems are easily expanded." *Transmission and Distribution*, September 1973.

"Supervisory Control," *The Lenkurt Demodulator*, February 1968.

"Supervisory Control, 51 Series," British Columbia, Canada: GTE Lenkurt Electric of Canada LTD., 1973.

"Third Generation Supervisory Control Systems," British Columbia, Canada: GTE Lenkurt Electric of Canada LTD., May 1973.

GTE LENKURT

DEMODULATOR

DECEMBER 1973

computers and supervisory control

Supervisory and control systems are used to perform such services as data collecting, monitoring, controlling, and alarm reporting of unattended remote equipment. The incorporation of a minicomputer in an overall system design imparts considerable expediency and reliability to these functions.

Supervisory and control systems run the gamut from simple status reporting, to remote transmission of complex analog signals. Telecontrol systems transmit control commands to unattended remote stations to activate relays, operate valves, and perform many other forms of control functions. Status reporting systems continuously monitor the status of isolated sites. Alarm reporting systems report the occurrence of any unusual event that may be detrimental to the operation of an unattended site. Telemetery systems relay information such as measurements of pipeline pressures, flow rates, task levels, and electrical quantities. Where a combination of several systems is necessary, the use of a minicomputer to process the information can save time and expense.

Minicomputers in Supervisory Control

Supervisory control systems deal with many forms of discrete information, including digitized analog, which must be presented to the operator in some understandable format. Traditionally, large numbers of display lamps and pushbuttons have been used for this man/machine interface with digital data, while meters or recorders have been used for the interface with analog data.

As systems grow in size, display lamps become inadequate as the only means of monitoring the status of a system. Not only does the quantity of lamps become unwieldy, but the information conveyed by the lamps may become difficult to comprehend. What is needed is a means of gathering the data, storing and editing it, and providing as an output, an indication or decision based on the useful information contained in the data. Since a minicomputer deals with binary digits, and can store and process data and make logic decisions, it is ideally suited to applications in supervisory control systems.

A minicomputer associated with a supervisory control system is normally arranged to include the following functions:

(a) data requests – requesting data from a remote site
(b) data storage – storage of raw data received
(c) data logging – printout of status changes and periodic or requested lists of current status or value
(d) data processing – manipulations and calculations with raw data to produce useful outputs
(e) decision making – logic processes on raw data to produce a recommended course of action or, in closed loop control, to initiate action via the telecontrol system.

These functions are what distinguish the minicomputer-based supervisory control system from the earlier data-logging systems, and what add a new dimension to the techniques of supervisory control.

Third generation minicomputer systems for supervisory control have incorporated the experience gained in custom designed systems, and have become easier to use, while at the same time providing the exact information desired by the operations personnel.

Some of the important aspects of the new systems include:

(a) ease of operation in normal day-to-day use, with minimum training required
(b) readout of system conditions in clear and unambiguous terms
(c) rapid pinpointing of alarm conditions
(d) telecontrol commands that are simple to initiate, yet secure and free of possible errors that may occur in dialing or pressing of buttons
(e) ability to provide permanent records for future analysis without creating too much paper printout (a "paper mill" effect)
(f) reliable, low maintenance operation
(g) security against false operation
(h) reasonable initial cost and low expansion cost.

In a computerized system, these objectives are brought about through proper arrangement of computer hardware, choice of peripheral equipment, and application of new technologies in supervisory hardware. In addition, software plays a key role in accomplishing all of the above objectives. Figure 1 shows a diagram of a GTE Lenkurt 51H central processor designed to operate with a combination of supervisory and control systems.

Computer Software

The software portion of a computerized supervisory control system is often thought of as an "add-on" item which is added last to make the system go. Software is actually a very important part of the system, and could even be considered as the essence of the system with the hardware portion being only a means of executing the functions contained in the software.

The software for a supervisory control system is often provided as one or both of two separate programs. One is the real-time operating program which is always needed for the functioning of the system. It is responsible for all aspects of data acquisition, storage, processing, calculations, outputs, and internal hardware monitoring. It may also include tasks for encoding, decoding, bit checking, and other data transmission functions which would otherwise be done in hardware. The other program, which is optional, is the system configurator that allows the user to add station and point names, logic equations, and mathematical routines to the software package.

ATS Executive Program

A major advantage of a computerized system is its flexibility in allowing expansion or changes with a minimum of disruption and expense. To achieve maximum flexibility, the real-time operating program used by GTE Lenkurt is structured on a modular basis. It is based on an asynchronous tasking supervisor (ATS) executive program which consists of an interrupt supervisor and a master scheduler to allocate the computer's resources among several tasks. The interrupt supervisor diverts the computer from the main program because of a particular event or set of circumstances, to a specific address which is directly related to the type of inter-

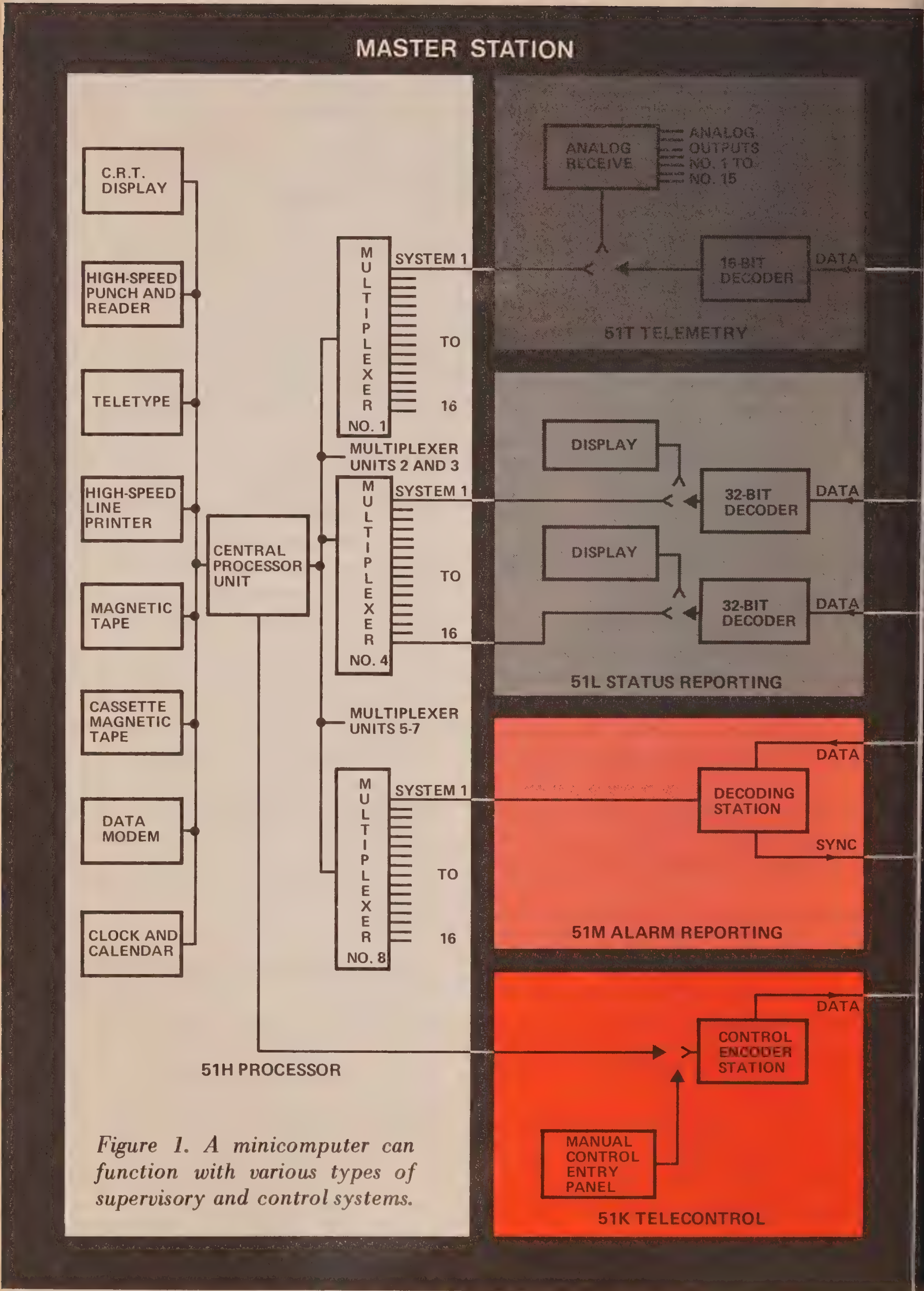
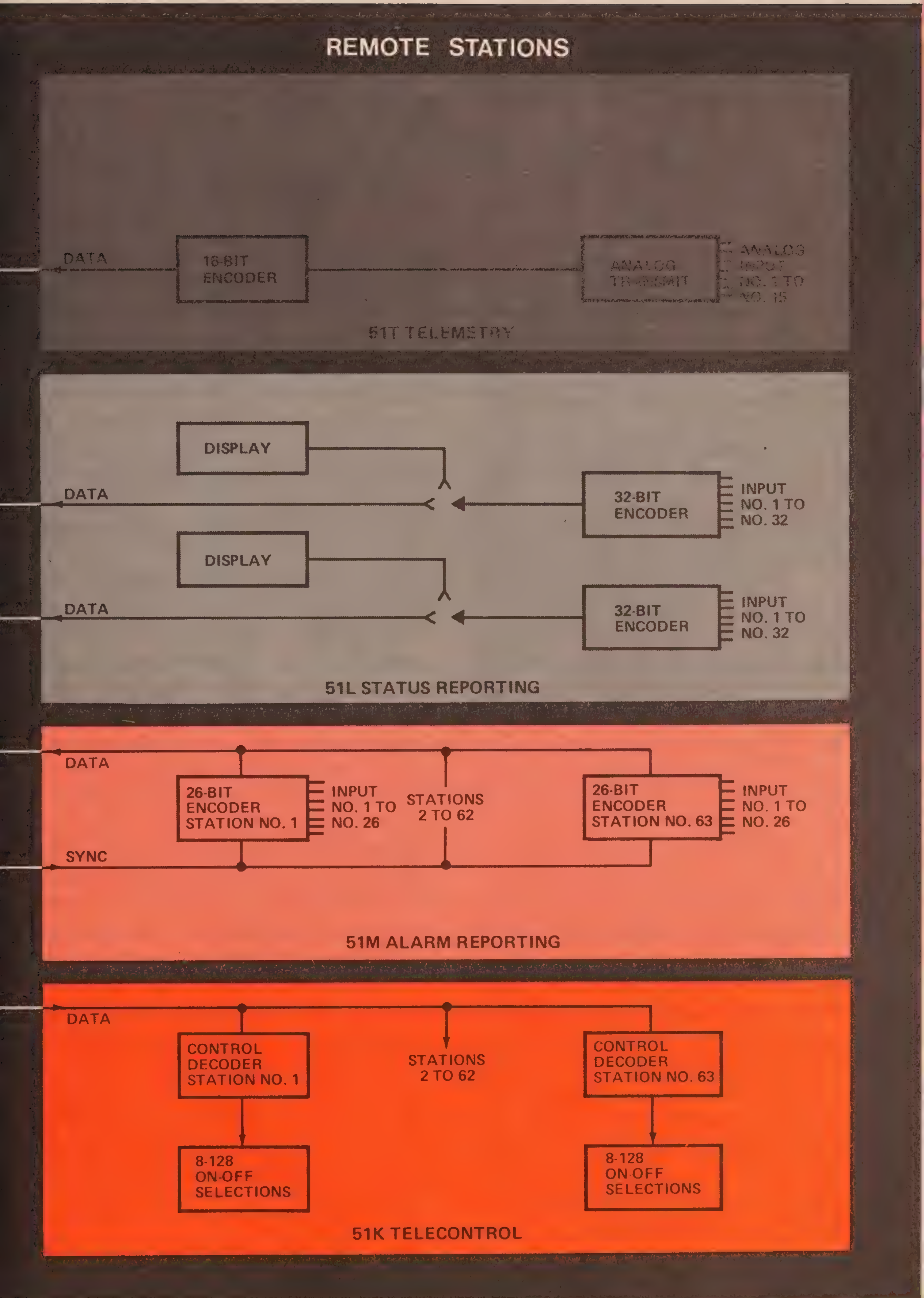


Figure 1. A minicomputer can function with various types of supervisory and control systems.



REMOTE STATIONS
DATA
16-BIT ENCODER
ANALOG
ANALOG
NO. 15
51T TELEMETRY
DISPLAY
DATA
32-BIT ENCODER
INPUT NO. 1 TO NO. 32
DISPLAY
DATA
32-BIT ENCODER
INPUT NO. 1 TO NO. 32
51L STATUS REPORTING
DATA
26-BIT ENCODER STATION NO. 1
INPUT NO. 1 TO NO. 26
STATIONS 2 TO 62
26-BIT ENCODER STATION NO. 63
INPUT NO. 1 TO NO. 26
SYNC
51M ALARM REPORTING
DATA
CONTROL DECODER STATION NO. 1
STATIONS 2 TO 62
CONTROL DECODER STATION NO. 63
8-128 ON-OFF SELECTIONS
8-128 ON-OFF SELECTIONS
51K TELECONTROL

rupt that has occurred. The master scheduler is called at periodic intervals to determine which program in memory is to be run. The tasks are programs, each of which is responsible for executing a certain function in the system. Typical tasks are timing, scanning, responding to keyboard requests, calculations, control and printout. The real-time operating program is comprised of all the tasks required by a particular system configuration, plus the ATS executive which supervises these tasks. To expand or modify the program, individual tasks are simply added or deleted.

The real-time operating program, due to timing and memory space considerations, generally accesses information in a coded format, and identifies station and point locations by a numerical address. For example, stations may be identified as station #00 to station #63, points from #000 to #127, and the point status as either a "1" or a "0". These numerical identifications can be translated to English language names and point conditions by using files produced by the user via the optional system configurator program.

System Configurator Program

The system configurator program is a compiler that permits the user to create a system definition file containing names of stations and points, phrases which appear in the printout, alarm limits for analog values, logic and mathematic statements, and other items which may be required in a specific system. The value of the system configurator lies in that the user can tailor the supervisory control system to his exact requirements and, because the program allows relatively simple inputs from the user, he does not have to be a computer expert to do this. Allowing the user to retain mastery over the computer in this way is extremely valuable, as he is the person most familiar with the operation of the supervised system and the requirements for supervision and control.

The logic and mathematical statements that can be put into the system by the user are really what distinguishes the computerized system from previous forms of supervisory control. In alarm reporting, statements can be added such that logic decisions must be made by the computer when certain combinations of alarms are received, and the output becomes a summary or recommended course of action rather than multiple alarm reports. Eventual closed loop control may follow from this system configuration.

In telecontrol, similar logic statements can inhibit undesirable commands and notify the operator of the reasons, such as attempts to shut down a main system when the standby system is out of service, or attempting to start machinery when a door alarm has been received (indicating a visitor on site). Usually an override command is available for use at the discretion of the operator. Mathematical statements can be added for telemetry measurements for scaling, setting alarm limits, or making calculations with the received values.

A very useful feature to have in the program is the ability to disable certain alarms via requests through the keyboard. This applies to cases where an intermittent alarm has been acknowledged or maintenance is being done at a site and the series of alarms received would be distracting to the operator.

Memory Space Requirements

The computers used in supervisory control are normally of the minicom-

OPERATING PROGRAM	ATS EXECUTIVE PROGRAM	0.5k WORDS
	TASK PROGRAMS	5.5k
	SYSTEM DEFINITION FILE	2.0k
	STATUS TABLE (POINTS X STATIONS)	4.0k
	QUEUE SPACE	4.0k
	TOTAL	16.0k WORDS

Figure 2. Typical storage capacity for a computerized supervisory and control system.

puter type, with core memory storage capacity of up to 32k words, expandable in about 4k or 8k word increments. The minimum storage required for a particular system is the sum of that required for the operating program, the system definition file, status tables, and output queue space. In a large supervisory system of 250 stations, each with 200 points, the required storage capacity would typically be as shown in Figure 2, based on the computer using 16 bits per word.

The queue space provides temporary storage for internal routines and allows outputs to be stored ready for printing, while the teleprinter is in the process of printing previously received messages. The amount of queue space determines how far the printer may fall behind the actual status of the system (without losing any of the data) when a greater-than-normal number of changes is received. Usually, any additional core memory which may remain is allotted to queue space.

Man/Machine Interfaces

Peripheral equipment is a subject much closer to the user than is the actual computer equipment, and also is more loosely defined in terms of actual devices and their arrangements. Since these devices perform the man/machine interface to the whole supervisory control system, considerable thought should be given to this equipment group by both application engineers and users.

Smaller systems can use teletype machines for hard copy printout and keyboard access to the minicomputer. These machines provide economical and reasonably reliable service, and have been available for a number of years. The use of high-speed line printers with larger systems increased the printing rate from 10 characters per second to about 350 characters per second, so the printout normally could keep up with status changes in the system. The "paper mill" problem, however, is even greater with line printers, since 350 characters per second represents about 250 lines or 10 pages per minute, and while much of the printout may be unnecessary in hard copy form for permanent record, it may be required for temporary indication, if no other means of output is available.

Today, the man/machine interface is often made through the combined use of CRT terminals, keyboards, hard-copy printers, and recording devices. The type of man/machine com-

munications handled by these devices include such functions as:

(a) alarm indications from a remote station
(b) status change indications from a remote station
(c) status of a group of points requested by the operator
(d) telecontrol command entry and check-back
(e) indication of selection and execution of telecontrol functions
(f) alarm lockout by operator for recurrent alarms or during maintenance
(g) indication of failures in supervisory control system
(h) data and time indication with each output
(i) results of logic evaluation of alarms status changes, or requests
(j) readout of telemetry values
(k) request for telemetry readout.

Depending upon the type of system to be supervised, whether it is a telecommunications network, power distribution system, oil pipeline, or transportation system, many of the above man/machine communications need not be printed out in hard copy. Others may require periodic printout, while a few require a permanent record of every occurrence.

Computerization should be considered today for any system whose total points exceed four or five hundred. Such a system might be a supervisory control system monitoring an electrical distribution network consisting of 15 substations, each with about 50 indications and 20 control functions. System totals of 750 indications and 300 control functions are more than sufficient to justify a computerized system. A full system status printout for 750 points would take at least 25 pages in hard copy, and would be largely useless, after once informing the operator of the system status. On the other hand, a CRT terminal can provide the same service without the paper, in less time, and with less effort from the operator. While the operator may not be able to read the CRT display any faster than he can a printed page, it *is* visible directly in front of the operator (no paper rolls or typing mechanisms in the way), and is associated with the keyboard from where the request came. Therefore, the reaction time is decreased, since the operator can begin to read immediately after requesting the information display, and can quickly erase the information and ask for a new display as soon as he has verified the status conditions.

When hard copy or other permanent record is required, either periodically or for selected points, it can be made by manual request through the keyboard or by a scheduling task in the computer software. In many cases, records are required for all alarm conditions, changes of status, exceeded preset limits, or control command inputs, and this can be scheduled by the software also.

Minicomputers may be used to activate lamps in display panels or mimic boards (single-line diagrams, displayed on the face of a control panel, representing the main connections of a system), and read control switches, where this remains the best means of operator interface, from a human engineering standpoint.

Backup Arrangements

In the interests of system reliability, all equipment on which the system is 100% dependent should be protected by duplicate on-line equipment or off-line standby equipment. System reaction times govern whether it is most practical to have a full time on-line standby or to allow a cold off-line minicomputer to update itself

when switched into service or to have no standby at all. Experience has shown that for some computerized supervisory control systems, duplicate on-line equipment provides many advantages in hardware economy, reliability, and operational features over standby off-line equipment. In the duplicate on-line arrangement, both minicomputers are hardwired to the channel multiplexing equipment and are normally continuously on-line and capable of performing identical functions. In the standby arrangement, one minicomputer is on-line, being connected to the channel multiplexing equipment through a switch, and carries out all functions required of the computer equipment, while the standby minicomputer is idle and can perform system operations only when it is switched on-line.

To provide back-up protection to each other, each minicomputer must be capable of wholly carrying the system. Therefore, computer hardware and software will probably be identical in both. In the duplicate on-line arrangement, the need for switching hardware is eliminated, and a source of possible unreliability is avoided. Switching between the two minicomputers for output is done internally in software, and adds another degree of flexibility to system usage, since software can usually be modified to suit new system requirements, while hardware is much more difficult and expensive to change.

Trouble Shooting Programs

One of the problems in maintaining a computerized supervisory control system is determining what has gone wrong when the system malfunctions. The trouble may be in the supervisory hardware, the computer hardware, or the software. To add to the problem, the maintenance man may only be familiar with the supervisory hardware. To assist in trouble shooting (and debugging), programs are available for hardware simulation and diagnosis.

Hardware simulation programs allow the operator to simulate the action of a hardware item from his keyboard, so that the minicomputer can respond to the inputs in the same manner that it would to the hardware item. These programs provide substitutions for hardware suspected of malfunction, as well as providing a check on correct software operation.

Diagnostic programs perform a series of tests on hardware items suspected of malfunction, and provide printouts whenever incorrect responses are received from the hardware. Together with the simulator programs, these programs enable the operator to pinpoint a trouble spot and restore service by substitution of equipment.

Supervisory Computer Systems

Typical systems consist of a central processor unit (CPU) and a number of interface units and peripheral devices needed for the particular application, along with a software package tailored to the application. In systems such as the GTE Lenkurt 51H supervisory computer system, a wide range of input and output services is available, including standard teletypes, medium and high speed printers, CRT displays, magnetic or paper tape devices, standard keyboards and touch-tone type entry panels (see Figure 3).

The programming, or software, portion of the system is structured on a modular basis, consisting of small individual programs for various tasks and an ATS program to supervise these tasks. This approach allows a software package to be tailored to a particular application in minimum time and at low cost, and has the added advantage

Figure 3. Operator area of a computer supervisory system.

of allowing changes to be made later for system expansion, change in operating procedures, or any other reason, without rewriting the complete program.

Flexibility in system arrangements is desirable. A computerized supervisory system may, for example, be arranged to operate unprotected, or it may be a fully protected system, having lamp displays, standby equipment, or duplicate on-line equipment as protection. More than one computer center may supervise the same operating system, where this arrangement is found advantageous.

Conclusions

Ten years ago a simple major/minor alarm from one unattended site was considered adequate — but not so today. Maintenance personnel and management alike, when receiving a call in the middle of the night, want to know exactly what is wrong before deciding to call out a repair crew. As a result, the demand for additional information from each location is increasing, with now an average of 24 to 32 separate alarm indications per site being required. High security remote control of such sites is also becoming a must, and often telemetering of certain operating measurements is desirable to further evaluate a problem.

Many alarm systems are initially being installed on a manually operated basis, utilizing visual status displays, with the knowledge that as the number of systems grows, a centralized minicomputer display system can be added economically.

As has been pioneered by oil and gas companies and power utilities, the operators of large communication systems, for example, are beginning to centralize alarm status reporting into one location, where immediate decisions can be made on utilization of alternative facilities and routes, as well as decisions on dispatching of repair crews.

Centralization of all information at one location, however, produces another problem: so much information may arrive at one time that the maintenance staff does not know what to attend to first. There is also the task of reading the alarm lamps or printouts to see what the trouble is and to determine its seriousness. All this takes valuable time.

The solution to many of the above problems may be a minicomputer to coordinate the system. The minicomputer is programmed to recognize faults, to evaluate their seriousness, to print this information out, and to recommend specific instructions on what to do.

The economical provision of such minicomputer systems, which can be readily maintained and added to by local maintenance forces, is a keystone of the new third generation supervisory and control systems. Such systems are now in service in many countries, and are being used for reporting and controlling of unattended microwave repeaters, for pipeline monitoring and control, as well as for remotely controlling and monitoring generating stations and substations by the power utilities.

Many users of supervisory and control systems will find the use of a minicomputer beneficial to their overall operation, whether it is included in initial system design or installed during system expansion.

BIBLIOGRAPHY

Kocher, Christopher P., assistant editor. "Applications of Minicomputers, Part 6," *EE/systems engineering today*, (July 1973), 66-72.

Pearce, J. Gordon. "Variations on stored-program control," *Telephony*, (September 24, 1973), 38-43.

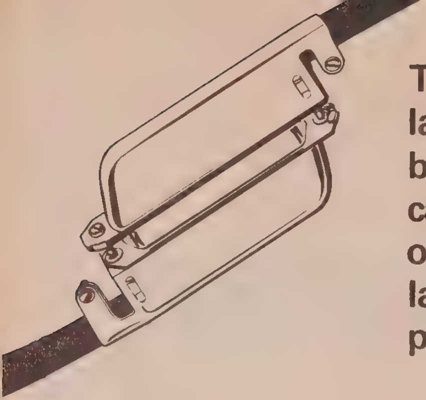
"Supervisory Control, 51 Series," British Columbia, Canada: GTE Lenkurt Electric of Canada, 1973.

"Third Generation Supervisory Systems," British Columbia, Canada: GTE Lenkurt Electric of Canada, May 1973.

GTE LENKURT

DEMODULATOR

JULY 1973



The tedious wiring procedure involved in the installation of new central office equipment has long been considered a way of life by the telecommunications industry. Connectorized equipment now offers a greatly reduced wiring task, a faster installation period, and a considerable reduction in the possibility of wiring errors.

The central office of a telephone company contains the necessary equipment to complete communications circuits between subscribers. Some of this equipment may include carrier, multiplex and microwave terminals, signaling equipment, jackfields, and other associated assemblies. Each piece of equipment requires a number of per-channel connections to other equipment, usually by way of a distributing frame. The distributing frame is a structure where the permanent wires of a central office are terminated; its function is to enable the placing of semi-permanent cross-connections to permanent equipment.

The necessary individual connections required to establish a working system may run into the thousands and will require many man-hours of installation and testing time. The possibility of wiring errors in such an installation is high, and valuable time is often wasted in tracing down the source of a wiring problem. Figure 1 shows a possible layout of a central office for one voice channel, as associated with a GTE Lenkurt 46A multiplex system; each "X" marks a point where there must be a connection to a distributing frame. If the number of connections for the voice channel shown in Figure 1 are multiplied by the hundreds of voice channels that typify an average central office, it becomes apparent that even the addition of a new terminal of equipment infers extensive time requirements in

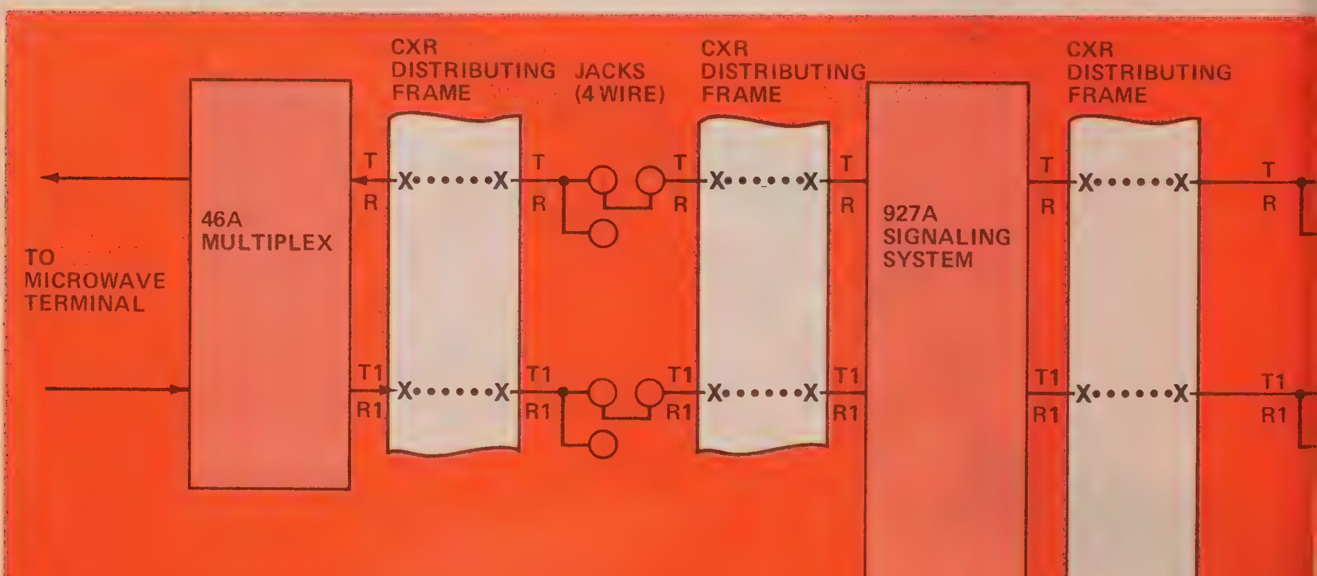


Figure 1. Excessive appearances at the distributing frame (marked by X) increase wire and installation costs.

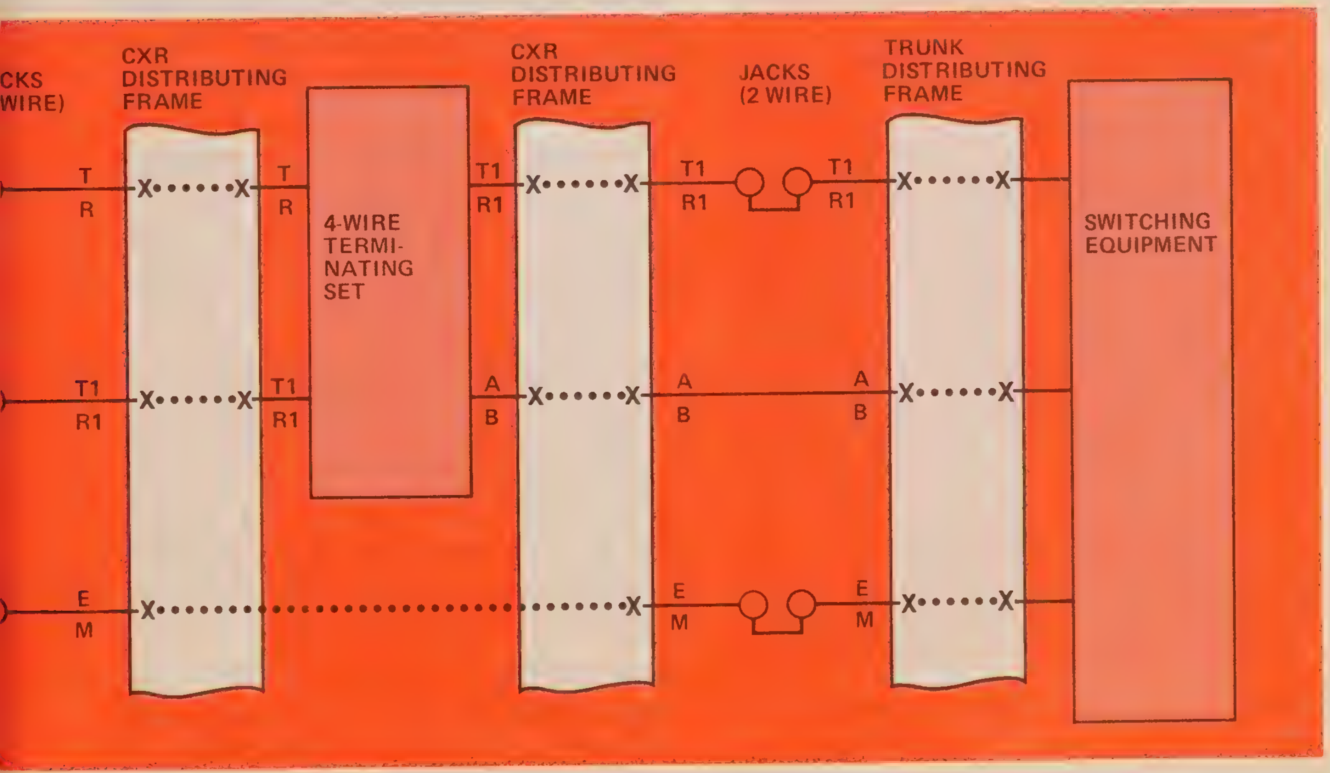
both installation and testing. The development of more sophisticated and compact telecommunications equipment means the presence of integrated circuits, transistors, and other complex and delicate components that can be destroyed as a result of a wiring error. The wrong voltage on the wrong pin may cause many hours of trouble shooting and wasted time.

Simplification – The First Attempts

One of the first methods used in an attempt to simplify the central office installation procedure was to provide cable stubs of an approximate length, prewired to carrier and multiplex channel equipment drop points. This method was further refined by the addition of terminal blocks at the end of the multipair cable stubs. The necessary length of the cable was specified by the customer and determined through measurements in the telephone office – from the carrier or multiplex equipment, up the racks, through the overhead ducts, and down to the terminations on the main distributing frame. The length of the cables – sometimes 200 feet – *did* present a shipping problem, since coils of cable had to be formed into large shipping crates alongside the equipment racks. When installed, slack in the cable presented a disordered appearance and took up excessive space in the overhead ducts. But this was still an improvement over manual on-site wiring.

Connectorization

The most recent development in the simplification of central office equipment installation is connectorization. The basic principle used by GTE Lenkurt in its connectorization schemes is relatively simple, and relies on the direction of connection. In a signaling system, for example, the line side of the equipment (the side toward the carrier) has female connectors; the drop side (the side toward the switch) has males. Also, pin assignments are standardized. Connector assignments are made by means of "key" telephone connectors (50 pin microribbon type) such as are commonly found in



business-office telephones. Such telephone instruments have six button keys to provide call holding, multiline pickup, signaling, and intercommunication. The term "connector" is often used to designate a female connecting device while "plug" designates a male connecting device. For simplicity, this article will use connector in the neuter sense, where it isn't essential that gender be specified. A key connector pin layout appears as shown in Figure 2. Each plug has 50 pins, with pins 25 and 50 generally left as spares. A twelve channel system requiring two wires per channel could, for example, be brought to the plug on the pins required for circuits 1 through 12. An additional 12-channel system could be placed on the remaining pins. Each channel has its own pair of colors, which remain consistent throughout the installation.

	CIRCUIT NUMBER	TIP	RING
GROUP 1	1	26	1
	2	27	2
	3	28	3
	4	29	4
	5	30	5
	6	31	6
	7	32	7
	8	33	8
	9	34	9
	10	35	10
	11	36	11
	12	37	12
GROUP 2	1	38	13
	2	39	14
	3	40	15
	4	41	16
	5	42	17
	6	43	18
	7	44	19
	8	45	20
	9	46	21
	10	47	22
	11	48	23
	12	49	24
	SPARE	50	25

Figure 2. One 50-pin connector can accommodate two 12-channel systems.

Figure 3 shows one channel of a fully connectorized GTE Lenkurt 46A3 multiplex system with 11A signaling. A comparison between the system shown in Figure 1 and that shown in Figure 3 reveals the need of far fewer interfaces for the connectorized system. This has been accomplished, in part, by the dedication of specific shelf positions to specific pins on the connector. The dotted lines in Figure 1 indicate jumpers or cross connections, which are semipermanent connections that may be changed for equipment rearrangement purposes. This means that any carrier channel can be assigned to any signaling equipment, to any 4-wire jack, etc.; there is any number of possible assignments.

Connectorization, along with universal signaling systems such as the GTE Lenkurt 11A, have made it possible to significantly reduce the number of appearances required at the distributing frame. The fully connectorized system of Figure 3 shows that a particular carrier channel is assigned to a particular set of jacks, which will be assigned to a particular terminal point on the 11A, which is assigned to another particular set of jacks, which is then cross-connectable only in terms of a trunk. Variations in circuit arrangement are accomplished in the connectorized system by changing interface types in the 11A signaling system, instead of changing cross-connections at the distributing frame, as is presently done.

Connectorization moves much of the installation chore out of the field and into the factory. Connections are made in the factory under ideal conditions where there are testing capabilities that permit testing of a complete system. Instead of requiring field connections of individual wires to many drop points on the equipment – in some cases as many as 10 leads per channel – connections are performed

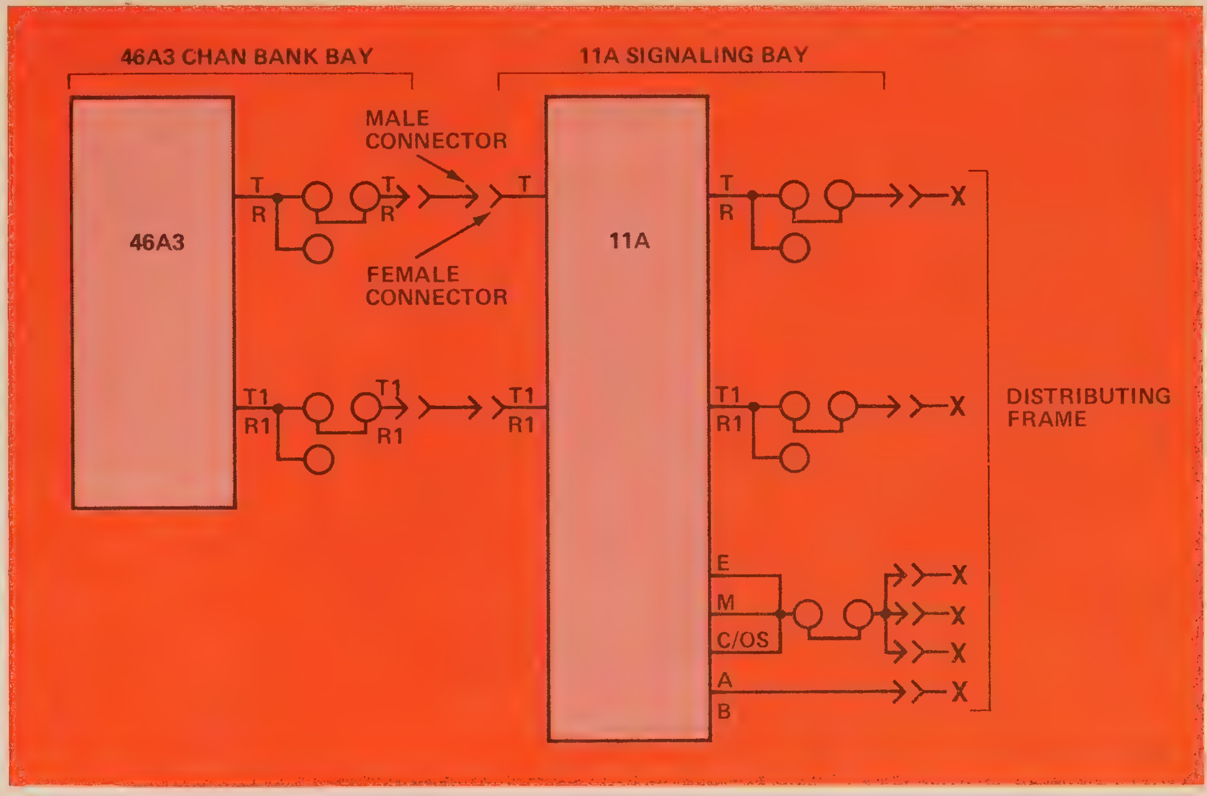


Figure 3. A fully connectorized system has a minimum of appearances at the distributing frame.

by snapping factory-wired, 50-pin connector sockets together. For this purpose,- jacketed cables in standard lengths of up to 300 feet are available with connectors on one or both ends, as required. In the field, most of the connections that need to be made are from a connector to the distributing frame. Figure 4 shows a connectorized rack of equipment ready for shipment.

Connectorization For PCM

PCM systems such as the GTE Lenkurt 9002A D2-type cable carrier have also been fully connectorized. Figure 5 shows the necessary leads per channel that must be connected on a D2-type system. The T and R leads are the transmit pair, T1 and R1 are the 4-wire receive or 2-wire transmit and receive pair, E and F are the receive signaling leads, M is the transmit signaling lead, C/OS is the trunk control lead, and A and B are the talking battery leads. Since these ten leads constitute one channel, the normal 24-channel PCM terminal will require that 240 connections be made for the full complement of equipment – a time-consuming task in the field, but avoidable through connectorization, and with the added benefit of factory testing to reduce the possibility of wiring errors.

More Advantages

Because connectorized equipment is essentially fully wired prior to shipment (a few HF, power, alarm and ground wires remain to be made in the field), the initial savings in installation and testing costs are significant. The ribbon type connectors used in connectorization are of proven reliability – there are more than 100 million in use today, and experience has shown that their failure rates are extremely low.

Wiring errors during initial equipment installation are a major factor responsible for premature failure of delicate electronic modules. Many of these errors can be eliminated by more accurate factory wiring procedures.

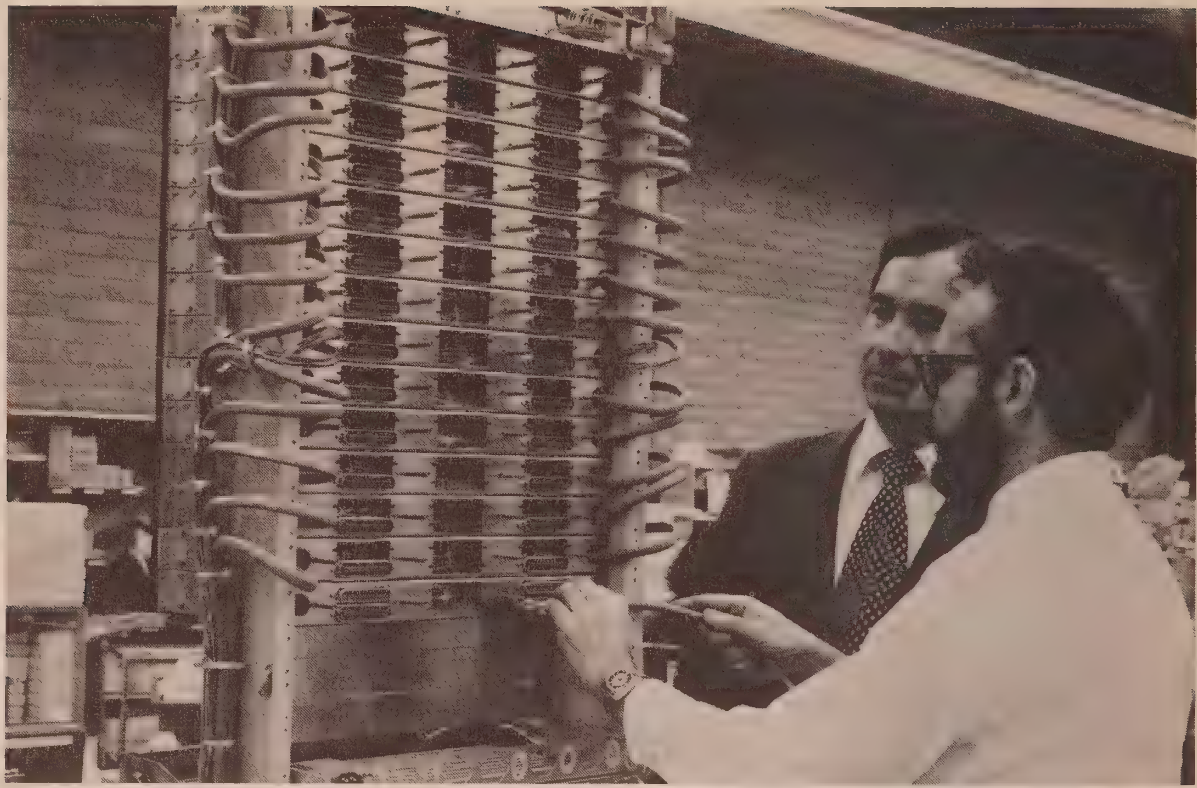


Figure 4. Much of the wire-by-wire connecting is eliminated by using connectorized terminals.

Even automatic testing techniques are being perfected that can reduce the wire-by-wire continuity checks of a typical 144-channel carrier equipment assembly from four hours to less than 15 minutes.

Another important use for connectorization is in the wiring of jackfields that are used for the patching and testing of telephone circuits (see Figure 6). In the field, this is usually a tedious and time-consuming wiring task, but factory connectorized jackfields significantly simplify the installation process. For example, in an equipment bay for up to 300 channels of type 46A multiplex equipment, complete with factory-installed jacks, all intra-bay channel-drop connections to the jackfield are completed at the factory, and external leads are terminated in connectors for extension by plug-ended cables (see Figure 6). If terminals blocks on the distributing frame are also equipped with connectors, on site installation procedures are reduced to the simple snapping together of connectors, once the cable is run through ducts. Prewiring from the distributing frame is possible with connectorization, and it is practical to do

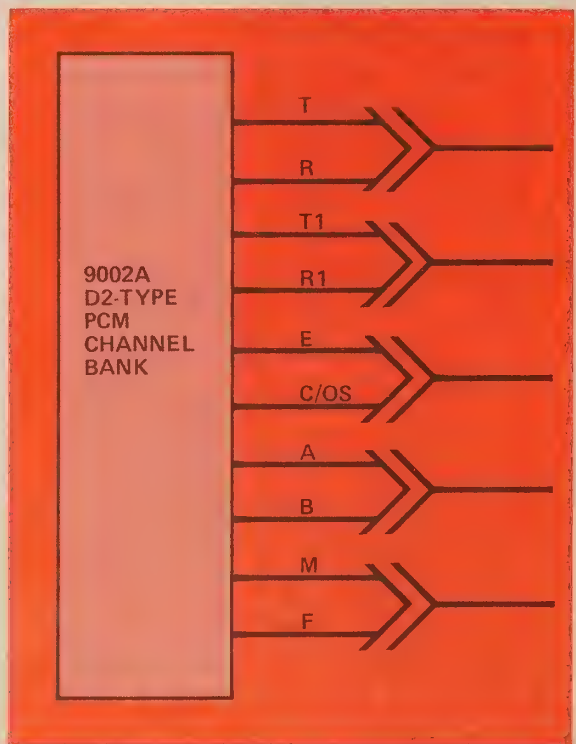


Figure 5. A D2-type PCM channel bank requires 10 connections per channel.

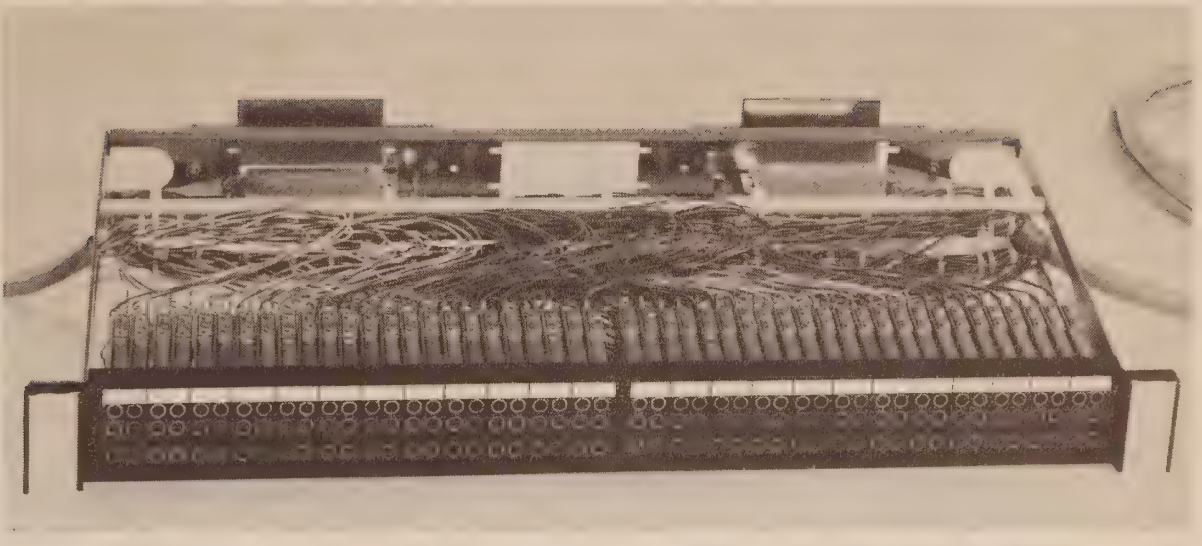


Figure 6. The task of jackfield wiring is moved from the field to the factory with connectorization.

much of the preliminary installation work in advance of receiving the carrier or multiplex equipment.

Jacks are used for testing a channel to determine if it is working properly and to reroute facilities when a temporary arrangement is necessary, such as when an important person, like a government official, for example, is briefly visiting an area. Public toll circuits must be attached to provide him with special circuits to the Capitol. Such temporary special service arrangements are usually made from the patching bays rather than by interconnection.

Lower Cable Costs

In lessening the number of trips to the distributing frame, the expense of actual physical cable is greatly reduced. A trip from a carrier terminal to a distributing frame may require from 50 to perhaps 200 feet of cable. The system in Figure 1 would require something in the neighborhood of 1500 pair feet of wire per circuit (if each trip to the distributing frame were 50 feet), while the connectorized system shown in Figure 3 would require only 300 pair feet of wire per circuit. This constitutes a reduction in cable requirements of 5:1, and implies an economic savings as well as conservation of space and a weight reduction in the overhead ducts.

The results of connectorization as a time-saving device are also impressive. Regular hard wiring time has been reduced an average of 75% through connectorization of telecommunications equipment. Where it once took about 24 man-hours to wire 48 circuits of miniature jacks in the field, with about one-fourth the time allotted to running and terminating cable at a main distributing frame, the process now consists of snapping and locking connectors, and running and terminating the cable as before – a time savings of 75 per cent.

Connectorization offers a new approach to equipment installation, which basically employs two concepts: (1) specific circuits in one piece of equipment are dedicated to specific circuits in associated equipment, (2) the dedication of specific circuits is realized and maintained by the use of connectored cables.

Connectorization is a logical and practical approach to telecommunications equipment installation. And, it can be applied to installation of new equipment in existing offices as well as to a completely new installation.

SECTION VI
DEVELOPMENTS AND DESIGN

GTE LENKURT

DEMODULATOR

NOVEMBER/DECEMBER 1975

OPTICAL COMMUNI-CATIONS

421

Also in this issue: **Lasers**

New technological advances in telecommunications often propose an alternate approach to doing the same thing, only better or more economically. While these new technologies may not totally and immediately replace previous ways of doing things, they often hint at what the telecommunications systems of the future may be like. Today the telecommunications industry stands on the threshold of a new communications concept -- the transmission of information by the use of light. Although the ideal optical communications system is far from a reality, telecommunications companies are busily developing and testing prototypes with that goal in mind.

What is light? What is its nature? Questions like these have stimulated the intense curiosity of the human mind for thousands of years. Ancient scholars had very little concrete knowledge of the nature of light. They surmised that light was composed of many particles emitting from a source; it was even conjectured that perhaps the eye itself emitted particles of light to illuminate objects. Surprisingly, these scholars did establish some theories about light that are still held today, including the idea that light travels in a straight line, that the reflection of light from a mirror is at an angle equal to that at which the light beam meets the mirror's surface, and that a beam of light is bent, or refracted, when it passes from air into a transparent material such as water or glass.

Early Experiments

Experiments conducted in 1666 by Issac Newton made great progress toward determining the nature of light. By noting the results of passing a beam of light through a prism, Newton concluded that white light was really a mixture of light components, each of which was capable of stimulating the eye in such a manner that it produced the sensation of color (see Figure 1). In a second experiment, Newton demonstrated that white light could be decomposed into its seven spectral colors by passing it through a first prism, then recomposed again by passing the dispersed light through an inverted second prism. Newton's experiments lent support to the popular theory that light was made up of tiny particles traveling at an extremely high speed, which would explain both the straight-line travel of light and refraction, since the particles would slow down when traveling through mediums denser than air. If, however, light indeed consisted of high-speed particles, some questions arose that remained unanswered. Why, for example, was one color of light refracted more than another; or why did the crossing of two beams of light not cause the streams of particles to collide, thereby distorting the individual paths of the beams?

In 1678, a Dutch physicist by the name of Christian Huyghens theorized that light was composed of waves whose varying lengths corresponded to different colors. This theory would explain the variation in refraction of different colors of light, since it was reasonable to assume that waves of different lengths would have varying degrees of refraction. From Huyghens'

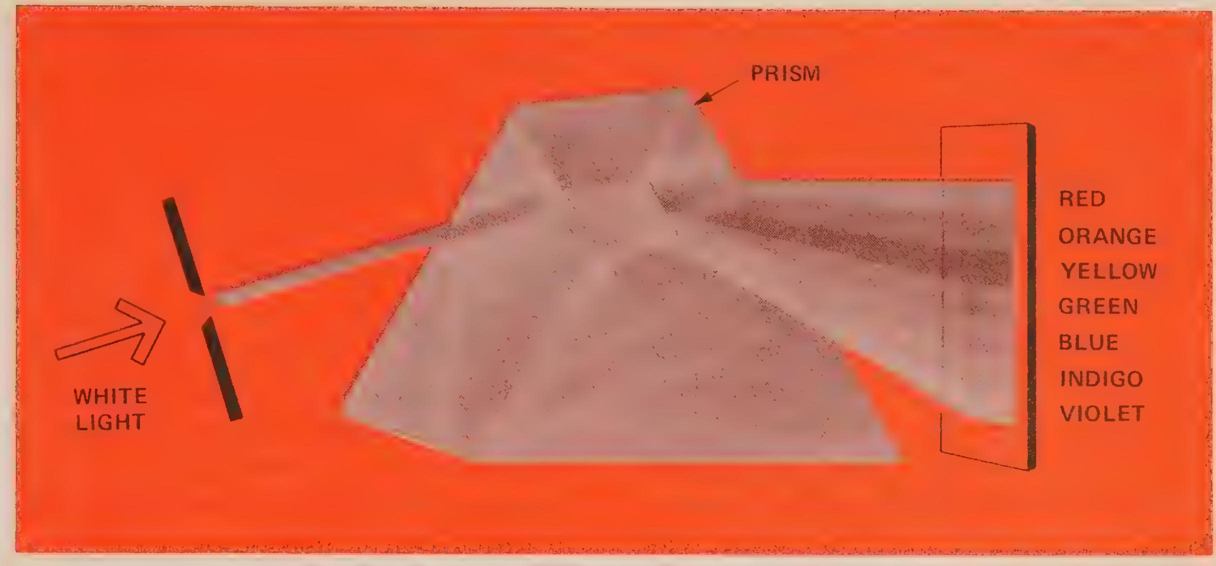


Figure 1. White light is a mixture of various light components. When white light is passed through a prism and projected on a screen, the components are dispersed according to color and wavelength.

theory, it could be explained that two beams of light projected from different directions and crossing each other did not become distorted by virtue of wave action, just as sound waves are able to cross without becoming distorted. In spite of being able to answer questions that could not be explained by the "particle theory," Huyghens' theory did not explain why light waves did not travel around objects as did sound and water waves, or how light waves could travel through a vacuum (the space between the sun and the earth, for example). Furthermore, if light consisted of waves, what was the medium being waved in outer space?

The answers to questions about light came slowly in spite of determined investigations. In 1818, a French physicist by the name of Augustin Jean Fresnel, whose concepts would later be extensively used in microwave communications, fortified the wave theory by showing that if an obstacle within a beam of light is small enough, light waves will definitely bend around it; the obstacle's size, however, must be close to the wavelength of light for this "diffraction" to take place. If an object which presents an obstacle to a beam of light is large with respect to the wavelength of light, the light not obstructed by the object will travel straight and cast a sharply-defined shadow: no diffraction takes place.

Along with attempts to find out more about the nature of light were studies attempting to determine the speed of light. The Italian astronomer and physicist Galileo Galilei (1564-1642) was the first to attempt light speed measurement, but although his idea of measuring light at increasingly greater distances was correct, he did not have the necessary mechanical devices to make an accurate measurement. More than 300 years later, the German-American physicist Albert Abraham Michelson was able to measure the speed of light in a vacuum and found it to be 186,284 miles per second. Still greater refinements in measurement techniques enabled scientists, in 1963, to determine the speed of light as being 186,281.7 miles per second (2.998×10^8 meters/sec).

Even while an increasing amount of information was gathered on light, some of the old questions still remained, particularly the question of

how light, if it indeed consisted of tiny waves, could travel through the vacuum of space. Was there an "ether" beyond the earth's atmosphere that enabled the passage of light from the sun and other stars? Many scientists thought so.

It was the concept of lines of force and strength of magnetic fields proposed by Michael Faraday, and subsequent mathematical derivations of these fields in the 1860's by James Clerk Maxwell, that supplied new insight into the nature of light. The relationships between electricity and magnetism described by Maxwell essentially implied that electric and magnetic fields must coexist; one cannot exist without the other. Further, this coexistence extended to changing fields, where a change in a magnetic field brought about a corresponding change in the electric field, and vice versa. This change phenomenon was what Maxwell termed electromagnetic radiation, an energy field that propagated outward in all directions in direct proportion to the number, or frequency, of the changes. Maxwell calculated that the velocity of an electromagnetic wave was equal to that of the speed of light, and he speculated that not only was visible light an electromagnetic radiation, but that it was only part of a greater spectrum, much of whose wavelengths were not visible to the eye.

In spite of the new theories and speculations, the question of the ether was not answered. Did it exist or didn't it? In 1900, German physicist Max Planck proposed that radiation consisted of discrete units which he called quanta. Planck theorized that radiation could be absorbed by a body only in multiples of quanta, and that the energy contained per quantum existed in inverse proportion to its wavelength. This latter theory implied that some colors of light would contain a greater degree of energy than others.

A New Concept

German-born Swiss physicist Albert Einstein verified the existence of Planck's quantum units while working out an explanation of the photoelectric effect. Extending the quantum theory, Einstein later proposed that light traveling through space did so in a quantum form which he termed the "photon." Here, then, was a step back to the particle theory of light. However, Einstein proposed that the photon had properties not only of the particle, but of the wave as well, and that depending on the prevailing conditions, either one group of properties or the other was exhibited. This theory now made it unnecessary that an ether exist through which light waves must travel, since they could travel through the vacuum of space due to their properties as particles.

Research into the nature of light has given birth to new terms. In the measurement of the length of light waves, for example, it has been found that the wavelength of red light is around .000075 centimeter. Because numbers such as these are difficult to work with, a more convenient unit called the Angstrom (Å) was adopted. One Angstrom unit equals one hundred-millionth of a centimeter. The previous measurement for the red wavelength (.000075 centimeter) thus corresponds to 7500 Angstrom units. Another unit that is used in connection with the measurement of light waves is the micron. This unit of measurement is equal to one millionth of a meter, or 10^4 Angstrom units. Violet light waves, for example, are in the .38-micron range.

Light Sources

The visible light frequency spectrum appears within the confines of a

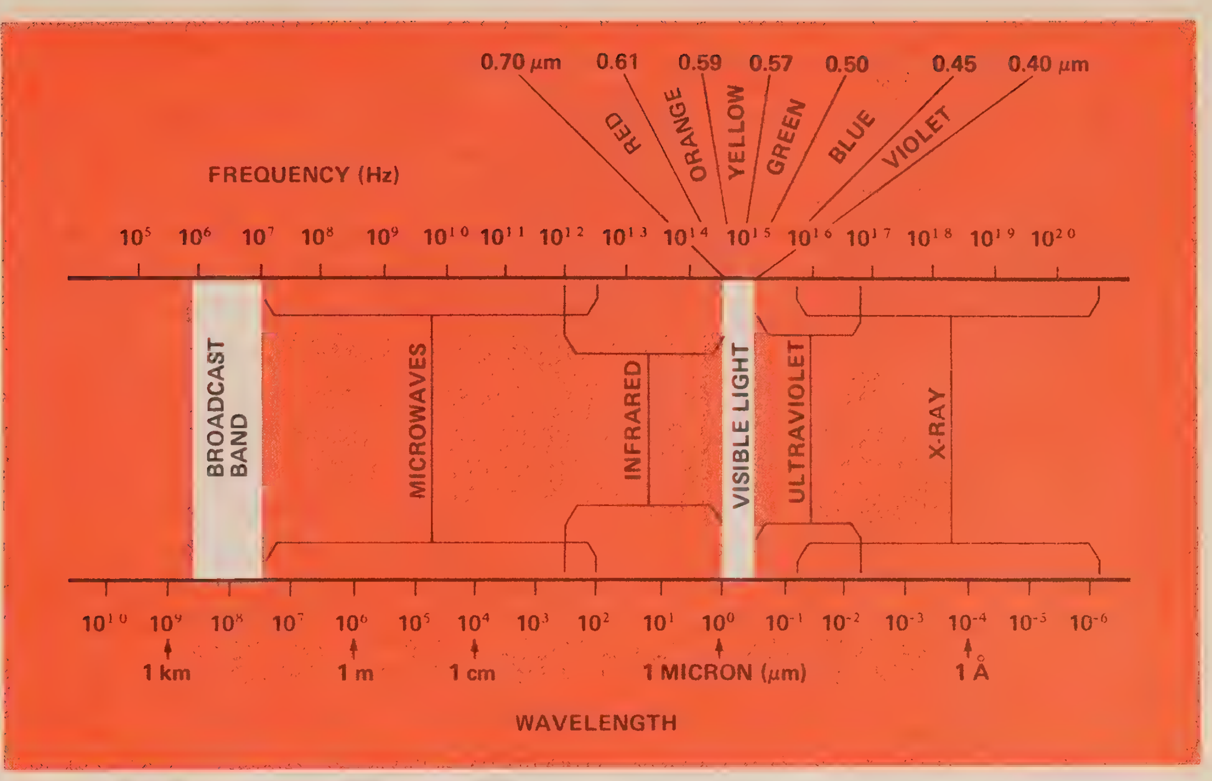


Figure 2. Under the title of light is included not only visible light, but also infrared, ultraviolet, and X-ray frequencies.

larger spectrum as shown in Figure 2. The immensity of the frequency spectrum of what is defined as light – which includes not only visible light but also infrared, ultraviolet and X-ray – has fascinated scientists with the possibility that it might be used to transmit information such as voice, radio, television, and data signals. Recent advances in semiconductor technology have produced two light sources that can be used in the transmission of light signals for communications purposes: the light emitting diode (LED) and the laser. At present, semiconductor lasers that can operate in the visible light spectrum at room temperatures for long periods of time are not yet commercially available. However, recent experimental results involving improvement in the growth of the crystalline material necessary in laser operation promises to bring long-life laser communications systems closer to commercial reality. Unlike the laser, the LED has been developed to the point where it is ready to take its place in experimental optical communications systems, although the ideal optical systems of the future will most likely use laser light sources for wideband communications.

The term "semiconductor" implies a material whose ability to conduct an electric current is limited. Charge carriers exist in semiconductor material, but since most of the electrons strongly adhere to the parent atoms, movement of charge carriers is restricted. To overcome, in a controlled manner, a semiconductor's resistance to electrical conduction, tiny amounts of certain impurities or dopants are added to the material. Dopant material which is composed of atoms having an excess number of electrons produces an "n-type" semiconductor, while one composed of atoms lacking electrons produces a "p-type" semiconductor. If slices of p- and n-type material are joined (a diode junction), the free electrons from the n-type material are combined over a thin portion of the junction (the depletion layer) with the

available holes in the p-type material. A voltage applied so that the p material is positive and the n material is negative forward biases the semiconductor, and causes current to flow. Polarity applied in the opposite direction (reverse biasing) causes current to cease flowing. In the case of an LED, current passing through a pn junction causes electrons to be temporarily "pumped" to a higher energy level, but as these electrons return to a more stable state, they release energy in the form of light of a certain color (or wavelength) which is dependent on the atomic structure of the semiconductor material. For example, an LED made of gallium arsenide (GaAs) material will emit light in the infrared portion of the frequency spectrum, while one made of gallium arsenide phosphide (GaAsP) will produce a visible red light. More technically, it can be said that light emission from an LED is caused by recombination of carriers injected across the pn junction.

The electroluminescent emission from an LED is multimode and incoherent light, which means that random relationships exist between the light waves emitted by the different atoms in the source. It is possible to produce a coherent light source, but this requires the use of a laser, which produces an intense light of a certain wavelength

Fiber Optics

Just as it is possible to send Morse code signals to a receiver some distance away by use of a flashlight or some other light source, so is it possible to send signals with an LED, but at a much faster rate and in far greater quantities. The LED typically generates 2.5 milliwatts of wide-angle infrared light in a wavelength band measuring from .87 μm (microns) to .92 μm. Light pulses of short duration suitable for use in T1, T2, and T3 PCM systems can be generated by electrically pulsing an LED.

An optical communications system requires, besides the light source, a medium over which the light signals are transmitted and a sensor, generally a semiconductor diode, which converts the light signals back to electrical signals. The medium by which light signals will be transmitted to a receiver in optical communications systems is most certain to be one of several types of hair-thin glass fibers, some of which are presently available for use and some of which are still in the development stage. An optical fiber is actually a tiny waveguide which supports optical frequency waves using the principles of total internal reflection at the boundaries of the fiber.

Transmission over optical fiber promises advantages over copper wire in the form of larger bandwidths, freedom from crosstalk and other types of interference, low cost, and light weight. In addition, much more information can be carried at optical frequencies than at the lower microwave frequencies. Bundles of optical fibers, each capable of carrying thousands of telephone conversations, could eventually replace miles of copper wire now being used for the same purpose.

Fiber Optic Modes

Fibers are usually categorized as one of three types: single-mode and multimode step-index fiber, and multimode graded-index fiber (see Figure 3). A single-mode fiber can function efficiently only by working in conjunction with the coherent light from a laser, since it has been determined that if the fiber's core is small enough, only the fundamental mode is guided along the fiber. A multimode fiber, on the other hand, may be used with incoherent light sources such as LED's. Figure 4 shows how incoherent light

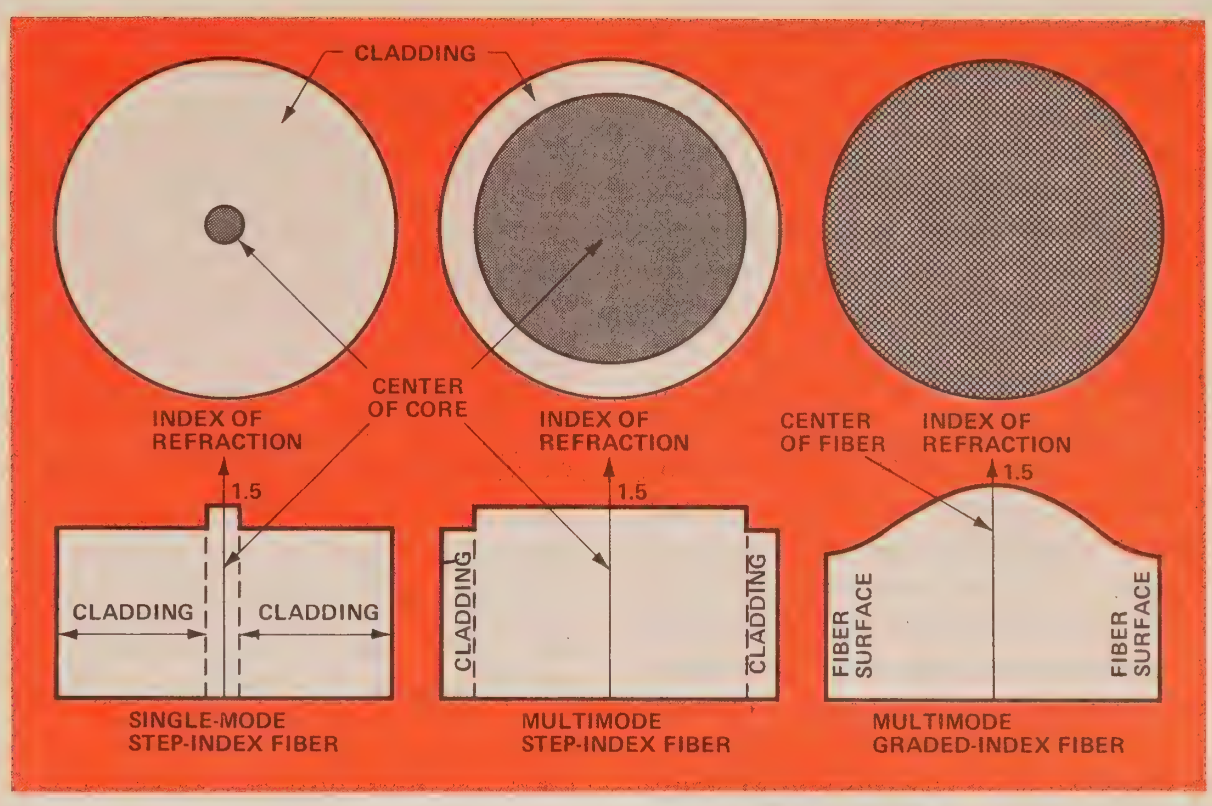


Figure 3. The three optical fibers most likely to be used for telecommunications are single-mode and multimode step-index, and graded-index multimode fiber.

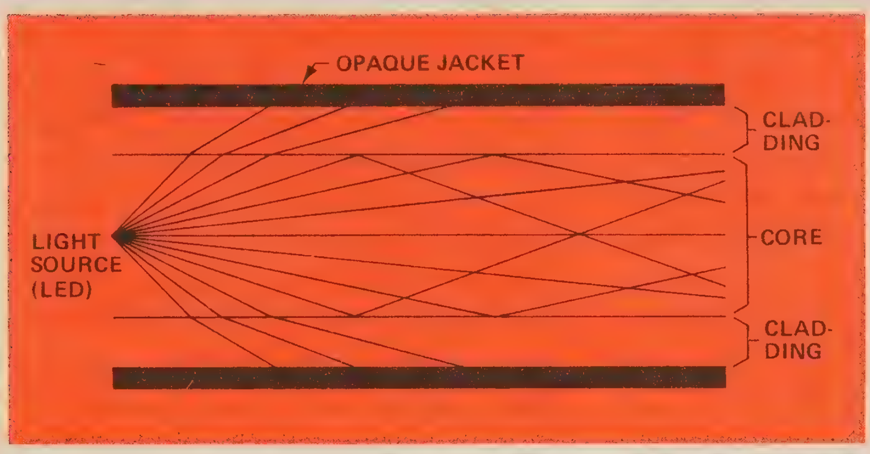


Figure 4. Incoherent light waves emitted from an LED are confined to the core of a multimode step-index fiber due to reflection from the cladding material. The light waves travel a zig-zag path along the fiber to the distant end.

waves from an LED travel along a multimode step-index fiber. Light rays are emitted uniformly by the LED from many points. Those rays not captured within the core of the fiber become totally absorbed by the jacket, while the others strike the interface area between the core and the cladding at angles which, by the process of total internal reflection, are forced to propagate within the boundaries of the core.

Because the core and cladding of an optical fiber are composed of materials differing slightly in index of refraction, the light from the LED travels through them at different speeds. The index of refraction of the core, being slightly higher than that of the cladding, causes light rays striking the interface at grazing angles to be reflected back into the core material; only a small amount of the total light propagates within the cladding region.

As in any communications system, the transmited signals in an optical fiber must span the distance to the receiver and arrive there in an accept-

able enough condition so that they can be detected with a certain degree of reliability. To this extent, the maximum range of a system largely depends on the type of light sources and light detectors used, and on the purity of the optical fiber and the nature of its construction.

Signal Degradation

Assuming that digital signals such as those produced by a PCM terminal were to be used in a hypothetical system, they would appear as bursts or flashes of light occurring within a uniform array of time slots. To be able to decipher the message that is represented by the signals, it is necessary that the receiving end be able to distinguish the bursts of light not only in intensity but also in time. Signal degradation in a fiber mainly occurs in the form of attenuation (dimming of light intensity) and differential delay (the broadening of the signal in time).

The extent to which a multimode fiber can accept and transmit light energy depends upon the angle at which the light rays enter the fiber. Relative to the axis of the fiber, this angle must be less than the critical acceptance angle (Θ_C) of the particular fiber being used (see Figure 5). In general, only about 4% of the total wide-angle light initially emitted by the LED is transmitted in the optical fiber. The attenuation of light energy traveling in a fiber is mainly due to absorption and scattering. Absorption loss is caused by the presence in the fiber of impurities such as iron, copper, nickel and cobalt. These materials usually are found trapped in the glass from which the optical fiber is made. For a good-quality fiber, the total amount of metallic-ion impurities should not be more than one part per million. To meet these requirements, intensive research by fiber manufacturers has produced fabrication

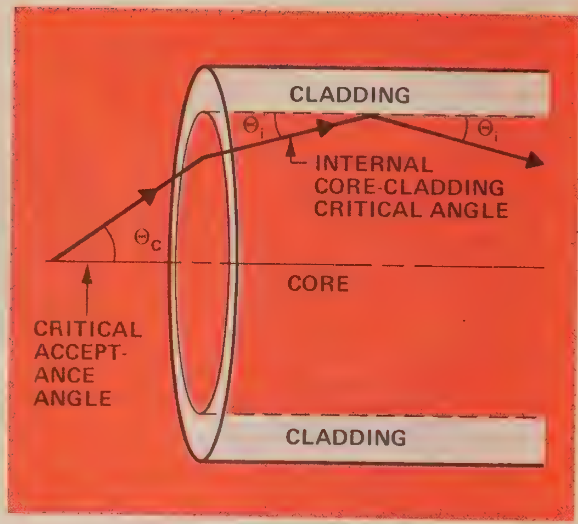


Figure 5. Light rays entering a step-index multimode fiber at greater than the critical angle (a higher order mode) will be absorbed by the opaque jacket.

methods that provide glass of such great purity that absorption loss is minimized in the manufacture of the latest experimental fibers.

Power loss due to scattering is caused by imperfections in the core material and by flaws in the region where the core interfaces with the cladding. Rayleigh scattering, another type of scattering which causes attenuation in optical transmission, is caused by the existence of tiny dielectric inconsistensies in the glass. Because these material inconsistencies are small with respect to the particular wavelength propagating in the fiber, scattering of light energy takes place in all directions, almost uniformly.

The transparency of optical fiber to be used for communications must be extremely high for a system to operate efficiently over an acceptable distance. For example, for a given intensity of light of a certain wavelength, a particular optical fiber might convey the energy a distance of 1000 meters, whereas in high quality optical glass and ordinary window glass or water, the energy would be conveyed only about 5 meters and 1 meter, respectively, for an equal amount of loss.

Differential Delay

The degradation of light by differential delay (pulse broadening or spreading) in modern optical fibers has more significant effects on transmission than does scattering. The cause of pulse broadening begins with the angle at which a ray from a light source enters the fiber. Those rays entering a multimode step-index fiber parallel to the fiber axis travel the shortest distance to the receiver, while those entering at various angles must be reflected by the cladding, and thereby travel a longer distance to the receiver. The difference in time of arrival at the receiver of the various rays causes a spreading of individual pulses. If the difference in arrival time between the fast and slow rays exceeds the time interval allowed between pulses, a pulse overlap occurs. Because pulse spreading increases with fiber length, it is important that light rays travel as close to the core as possible. For this to occur, the difference in refractive indexes of the core and cladding must be kept small, thus also keeping the critical acceptance angle small. Pulse broadening must be especially limited in systems processing higher bit rates, since higher data speeds mean a shorter time interval between pulses and, consequently, less tolerance for errors due to pulse spreading.

New Developments

Developments in fiber optic technology have come at a rapid pace. In 1970, Corning Glass Works developed a fiber with a loss component of 20 dB per kilometer; four years later, experimental fibers had been tested that yielded losses of only about 2 dB per km. How soon such fibers will be available commercially is difficult to predict, but such advances show that the concept of light transmission for communications will not fall short due to lack of the necessary technology.

A departure from the step-index method of confining light energy to the core is a graded-index optical fiber developed by the Nippon Sheet Glass Co. of Japan. Called Selfoc (abbreviation for "self-focusing"), this graded-index fiber consists of one material interspersed with a second material in such a way that the index of refraction decreases at a faster and faster rate with distance from the axis of the fiber. By this means, the light rays travel back and forth across the axis of the fiber in a sinusoidal manner, with the refractive index reaching a maximum value at the fiber's center and a minimum value at the surface (see Figure 6). Because the speed of a light ray varies inversely with the refractive index of the material through which it propagates, it will travel slower in areas close to the center and faster in regions farthest away from the center. The effect of this action is that all rays traveling in the Selfoc fiber will reach the receiver at nearly the same time.

One additional method of eliminating differential delay is by constructing a step-index fiber with a core so small that only a single electromagnetic mode is allowed to propagate. This single-mode technique eliminates the interference that is created when light rays of different wavelengths propagate along the fiber. However, this type of fiber construction requires use of the monochromatic light that only a laser can produce (see Figure 7). Also, the single-mode fiber is not only difficult to manufacture, but difficult to handle in practical applications, although fibers of such construction have the potential for carrying much more information than fibers of other design. Semiconductor lasers are currently being developed which may be used with single-mode fibers. This kind of compatibility between available components and those yet to be developed may some day revolutionize

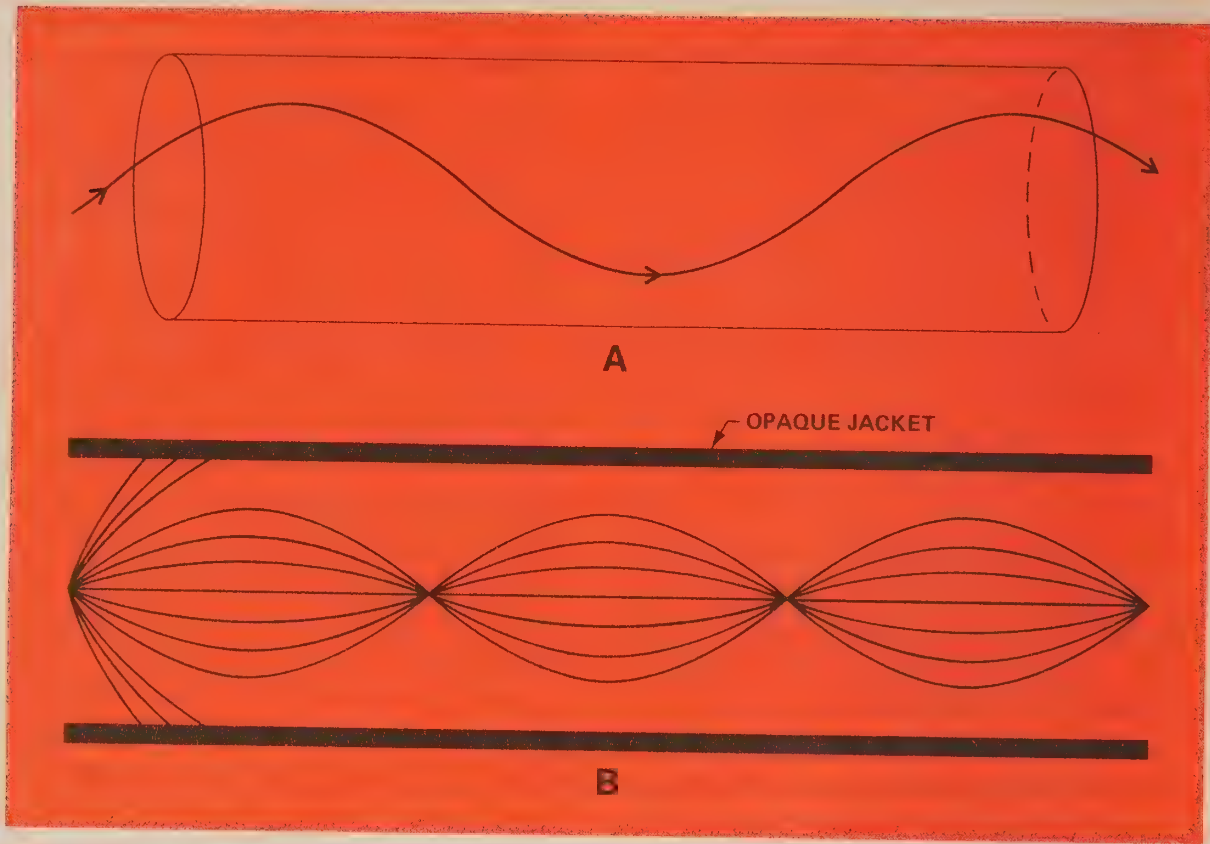


Figure 6A. A single ray travels a sinusoidal path along a graded-index fiber. The effect of the self-focusing property of graded-index fiber on incoherent light waves is shown in Figure 6B.

the field of wideband communications.

An Optical System

With the degree of technology presently existing in the field of electronics, it is theoretically possible to assemble available components into a working optical communications system. Indeed, this is one type of development that is being done in advanced telecommunications laboratories such as Bell Labs and GTE Labs. A simple experimental optical communications system might appear as shown in Figure 8. Signals from a D2-type channel bank such as the GTE Lenkurt 9002B are amplified and fed directly to an LED, which transforms the electrical signals to light and transmits them down the glass fiber at a T1 rate of 1.544 megabits per second. After the light energy in the fiber has undergone an attenuation of about 45 dB,

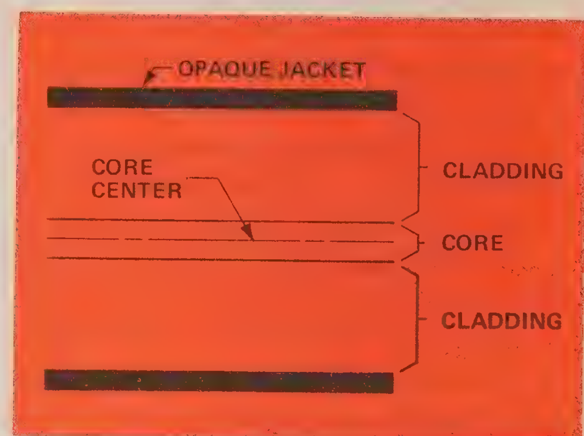


Figure 7. The step-index single-mode optical fiber must be used with the coherent light from a laser.

or when pulse distortion due to differential delay exceeds one-half of a time slot, it becomes necessary to regenerate the light pulses in a repeater so that they may be properly detected at the distant end. At the repeater, the degenerated light pulses first encounter a photodiode, which is a semi-

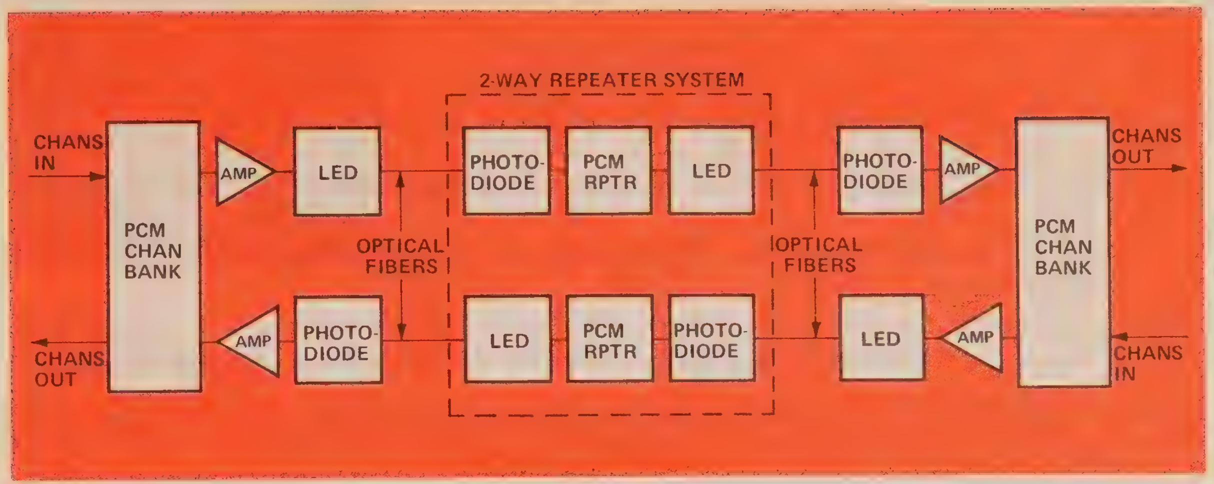


Figure 8. A simple, hypothetical optical communications system using currently available PCM equipment.

conductor device whose function is to convert light impulses into electrical energy. This electrical energy in the form of pulses is amplified, detected, timed, regenerated, and conveyed to an LED. The LED converts the electrical energy back to light pulses which are then sent into the following section of optical fiber. At the distant end of the line, the light pulses are again transformed into electrical pulses which are amplified and processed as usual in the PCM terminal. For signals transmitted in the other direction, the opposite action takes place.

The prospect of optical communications systems is truly attractive, since they not only promise to replace tons of expensive copper wire with hair-thin strands of silica, one of the most abundant materials on earth, but they also offer such potential advantages in future systems as bandwidth capacities of 50,000 voice channels or 30 TV channels per single fiber, using an LED light source.

Coupled with the advantages are, of course, technological problems yet to be overcome before these systems become a practical reality. Some of these include developing cables that will carry not only light signals, but also the electric power necessary for repeater operation. Adequate alignment and splicing techniques must be developed so that optical fibers may be serviced in the field, and still maintain their continuity, with a minimum of loss. Development of an all-optical repeater will also do much to maximize system length.

If technology progresses at its current pace, the not-too-distant future may hold such innovations as optical systems that use integrated optical circuitry (IOC) in the same way that today's electronic equipment uses integrated circuits, except that IOC components will consist of microscopic lasers, optical switches, and laser modulators.

To be sure, widespread use of commercial optical communications systems will depend on whether they are economically feasible. However the speculation on feasibility runs, research labs such as GTE Laboratories are planning to test prototype systems in the field this coming year. These tests will not only help determine the economics of optical systems, but also give some insight on additional developments that may be required for commercial system applications.

BIBLIOGRAPHY

1. "A Promising Light From Semiconductor Lasers," *Business Week*, September 15, 1975.

2. Asimov, Isaac. *Asimov's Guide To Science.* New York: Basic Books, Inc., 1972.

3. Besancon, Robert M., ed. *The Encyclopedia of Physics.* New York: Von Nostrand Reinhold Company, 1966.

4. Cook, J.S. "Communication by Optical Fiber," *Scientific American*, November 1973.

5. Free, John. "The Coming of Age of Fiber Optics: How Wires of Glass Will Revolutionize Communications," *Popular Science*, August 1975.

6. Fulenwider, John and George Killinger. "Optical T-Carrier Systems on Glass-Fiber Cable: A Promising New Technology," USITA Technical Session, Southeastern Telephone Showcase, Charlotte, N.C., 1975.

7. Fulenwider, John. "Some Potential Impacts of Optical Fiber Transmission and Integrated Optics on World Communications," *1975 IEEE Intercon Conference Record*, Session 5, April 8 – 10, 1975.

8. Haenichen, John C. "A New Era For Semiconductors," *The Electronic Engineer*, July 1970.

9. Li, Tingye. "Optical Transmission Research Moves Ahead," *Bell Laboratories Record*, September 1975.

10. Li, Tingye and Enrique A.J. Marcatili. "Research on Optical-Fiber Transmission," *Bell Laboratories Record*, December 1971.

11. "Optical-Fiber Communications Spurred by New Waveguide and Low Transmission Losses," *Electronics*, May 24, 1973.

12. Ramsey, Murray. "Beginning To Hear the Light," *Communications International*, October 1974.

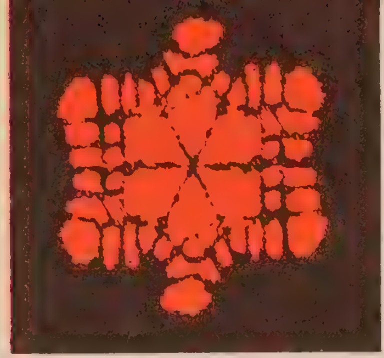
13. *Reference Data For Radio Engineers.* New York: Howard W. Sams & Co., Inc., 1968.

14. Rigterink, Merle D. "Better Glass Fibers for Optical Transmission," *Bell Laboratories Record*, September 1975.

15. Sussman, Aaron. *The Amateur Photographer's Handbook.* New York: Thomas Y. Crowell Company, 1965.

16. "The Light Wave of the Future," *Business Week*, September 1, 1975.

17. Uvarov, E.B. and D.R. Chapman. *The Penguin Dictionary of Science.* New York: Schocken Books, 1972.



Lasers

The laser has evolved from a science fiction concept into a working device which has found widespread use, from delicate eye operations and the treatment of skin cancer to a cutting source for intricate integrated circuits. In the telecommunications field, intensive investigation is being conducted in an effort to harness the enormous bandwidth of the laser beam for the transmission of information.

The laser (acronym for light amplification by stimulated emission of radiation) comes in a myriad of varieties, each specifically developed for a certain function. Laser types include: ruby, gas, chemical, liquid, metal-vapor, and semiconductor lasers.

The principle of the laser is based on atomic physics, and dates back to 1917 when Albert Einstein theorized that controlled radiation could be obtained from an atom (or molecule) under certain conditions. All matter consists of atoms made up of a "heavy" nucleus surrounded by electrons. Atoms in their natural state are usually in a relatively undisturbed, or "ground," condition because, normally, enough electrons surround the nucleus to neutralize its charge. That is, the energy of orbiting electrons is balanced by the energy in the atom's nucleus. In an atom's ground condition, electrons revolving around the positively charged nucleus are confined to a number of shells, usually designated K, L, M, etc. The maximum number of electrons allowable in the K shell is 2; 8 are allowed in the L shell, 18 in the M shell and 32 in the N shell. The electrons occupy specific orbits determined by their own energy levels. The distance of each orbit from the nucleus represents the specific amount of energy possessed by the electrons in that orbit. The closer the electron orbit is to the nucleus, the lower the electron energy level (see Figure 1).

Einstein suggested that "pumping" energy into atoms by means of external excitation would cause electrons to leave their natural orbits and rise to the second, third or higher level, depending on the quantity of energy applied. If enough energy is applied to the electron to raise it from one level to another, it will absorb only that amount of energy required for the jump. This pumping of energy places an atom in an "excited" state, and its natural reaction is to return to its ground state. When this takes place, a

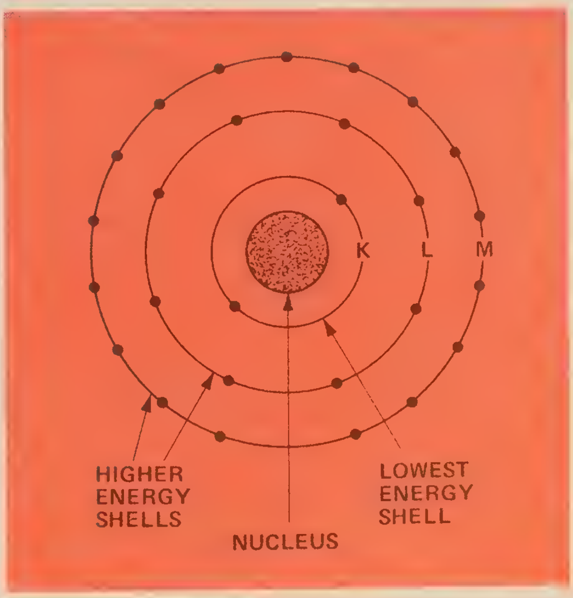


Figure 1. In an atom's natural state, electrons orbiting the nucleus are confined to a number of shells.

 ● Vol. 24 No. 12

photon of radiation is emitted. (A photon is a discrete packet of electromagnetic waves.) The energy of the photon is exactly proportional to its frequency — the higher the energy, the higher the frequency and the shorter the wavelength. For example, high-energy photons may appear as X-rays or ultraviolet radiation, while those of lower energy may give off visible light of any color, radiant heat, or radio waves. Also, the exact amount of energy absorbed or emitted by an electron in jumping from one energy level to another differs with each type of material and for each combination of electron shells. This means that an electron will emit radiation of one wavelength in dropping from the second to the first shell (the shell closest to the nucleus) in a substance, and a second wavelength when dropping from the third to the second shell. This can be observed in such familiar examples as neon light, where the atoms in molecules of gas are excited to upper energy levels by the presence of a high voltage. As the excited and ionized gas molecules drop back to their ground state, they emit light of a characteristic white color. This also accounts for the yellow-orange light of sodium vapor lamps, and the bluish-green from mercury vapor.

Optical Pumping

Pumping is a term generally used to mean the raising of matter from one energy level to a higher one. Light is a particular type of matter, and in optical pumping light energy is the power source used to raise individual atoms to higher internal energy states. A simplified explanation of the pumping process can be made with the help of Figure 2. The atoms in Figure 2 can be made to achieve any one of three energy levels, designated A, B, and C. Levels B and C are of low energy and are spaced very close together. The differences between the energy levels correspond to photons vibrating at certain frequencies. Before pumping, the atoms are distributed evenly between levels B and C (2-1). If this group of atoms is bombarded with a beam of light from which the spectral line AB has been removed by filtering, that light beam contains photons capable of exciting atoms at level C but not at level B. The atoms excited at level C absorb energy and rise to level A (2-2), where they will remain for a very short time (as short as one ten-millionth of a second), then return to either level B or C, emitting energy as they do so (2-3). Once an atom drops to level B, it can no longer be excited by the incident light. Given enough transitions between levels C and A, each atom will eventually arrive at the B level, which means the material has been totally pumped (2-4 and 2-5). Atoms can be returned to energy level C by irradiation at a frequency which corresponds to the transition energy between the C and B levels (2-6).

The Ammonia Molecule

The road to the development of the laser began with studies of the ammonia molecule (NH_3), which contains three hydrogen atoms and one nitrogen atom arranged in a pyramidal structure, and can be displayed in the manner shown in Figure 3. One hydrogen atom is positioned at each corner of the base of the pyramid, with the nitrogen atom occupying a position at the apex. The ammonia molecule can be made to vibrate by irradiating it with microwave energy, which means that the nitrogen atom is repeatedly caused to travel a path through the plane of the triangle to a corresponding position on the other side, and back again. This vibration is extremely consistent at 24 billion times a second (2.4×10^{10} Hz). In 1949, using this property of the ammonia molecule,

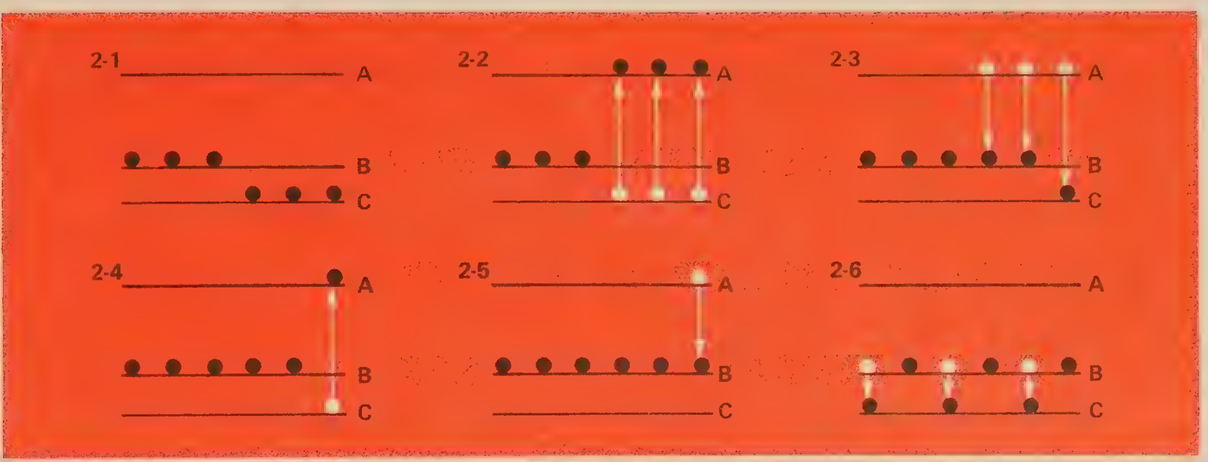


Figure 2. Optical pumping involves the transferring of atoms to a higher energy level by external excitation of the laser medium.

American physicist Harold Lyons constructed an extremely accurate atomic clock. By 1964, atomic clocks capable of such chronologic accuracy as one second of variation in 100,000 years were being produced.

A vibrating ammonia molecule emits electromagnetic radiation at 24 GHz, which corresponds to a microwave wavelength of 1.26 centimeters. The ammonia molecule may be seen as being able to occupy two discrete energy levels. The difference in energy between these two levels is equivalent to that of a photon providing a radiation of electromagnetic energy with a wavelength of 1.26 centimeter. A photon with these characteristics is emitted when an ammonia molecule falls from the higher energy level to the lower level. Similarly, when an ammonia molecule located in the lower energy level absorbs a photon with the appropriate characteristics, it ascends to the higher level.

When an ammonia molecule which is already located in the higher energy level is exposed to additional photons corresponding to a wavelength of 1.26 centimeter, the molecule will be forced back to the lower level, but will emit a photon of the exact size and traveling in the same direction as the entering photon. Essentially, this means that two identical photons exist where only one existed before. Molecules of ammonia bombarded by microwave frequency radiation can be pumped from a lower energy level to a higher level and vice versa. It occurred to scientists that if all the available ammonia molecules could be conveyed to the upper energy level, bombardment by a beam of radiation at a microwave frequency should have a

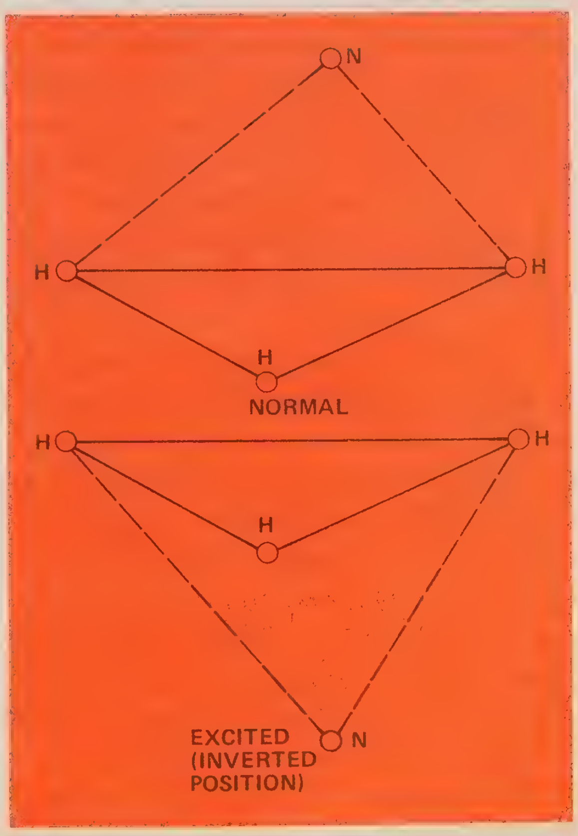


Figure 3. The ammonia molecule can be caused to vibrate by irradiating it with microwave energy.

cascading effect as far as release of energy was concerned. The microwave radiation would provide a photon which would strike an ammonia atom, thus forcing it to a lower energy level, but causing it to emit a second photon in the process. These two photons would then be capable of striking two more molecules, thus releasing two more photons, etc. In this way, the initial action of a single photon would create a deluge of new photons of equal frequency and direction.

The First Maser

By 1953, an American physicist by the name of Charles Townes had succeeded in isolating ammonia molecules occupying the high energy level. He then bombarded these molecules with photons of the proper wavelength, causing a stimulation which created the appearance of a great number of photons. In effect, this process signified a form of amplification, since the ratio between the number of input photons and the number at the output was very high. This was the first example of a gaseous "maser" (acronym for microwave amplification by stimulated emission of radiation). Masers were soon developed which used solid material rather than ammonia gas.

The first masers required pumping to the higher energy level before stimulation. This came to be called an intermittent maser, since after delivering a burst of radiation of very short duration, the pumping process had to be repeated before stimulation could occur again. The delay caused by having to pump and stimulate at discrete intervals led to the development of the 3-level system of operation.

In devising the 3-level system, Dutch-American physicist Nicolaas Bloembergen theorized that by selecting for the maser core material a substance containing electrons in each of three discrete energy levels, pumping and emission could take place at the same time. The eventual incorporation of this process led to the development of the continuous maser.

The basic principle of maser operation can be applied to electromagnetic waves of any length. When such principles are applied to electromagnetic waves in the visible light spectrum, the device may be called a laser.

The First Laser

In 1960, an American physicist by the name of Theodore Maiman succeeded in constructing the first laser. In Maiman's device, a bar of synthetic ruby (which is mainly aluminum oxide with traces of chromium oxide) was exposed to high-intensity light. This caused the electrons of the chromium atoms to be momentarily pumped to higher levels. As the electrons retreated back to the lower levels, the first photons which were produced collided with other atoms which in turn produced additional photons.

The essential elements of Maiman's ruby laser appear as shown in Figure 4. The laser ruby is polished to optical flatness and silver coated at each end. One end of the crystal, however, is heavily silvered so that it reflects all light, while the other end is more thinly coated, and reflects only about 92 percent of the light incident on it. Around the crystal is a helical xenon flash tube which provides the intense light necessary for optical pumping. When photons from the flash tube irradiate the ruby crystal rod, the energy of some of the chromium atoms rises from the ground state to higher energy levels. As the elevated chromium atoms drop back toward the ground state, some come to rest temporarily at an intermediate or metastable (only slightly stable) state. Because the flash tube continues to irradiate additional chromium atoms,

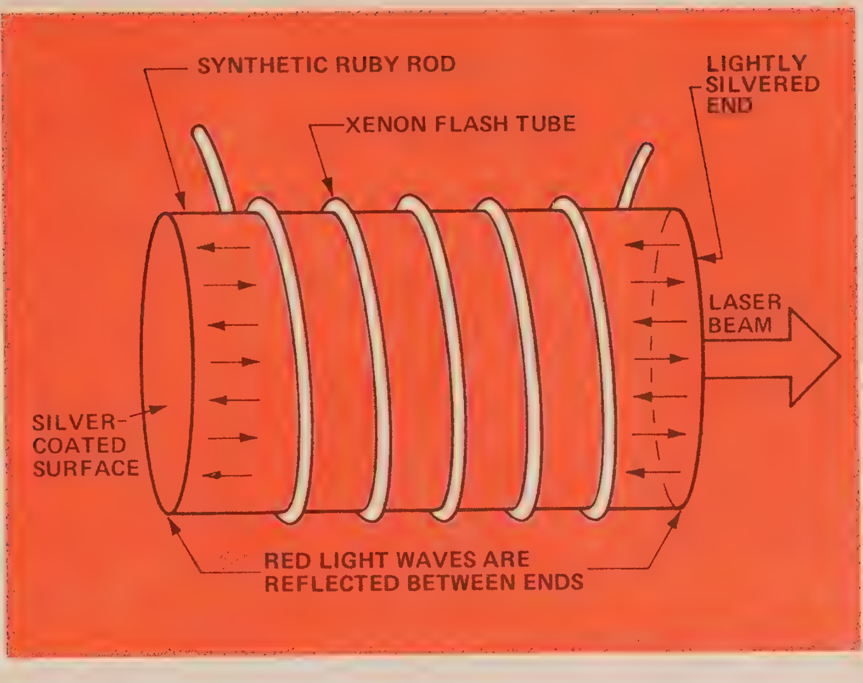


Figure 4. Theodore Maiman's first laser had silver coatings on the ends of the ruby rod; later adaptations of the device utilized external silver mirrors at the ends of the rods.

more of these atoms collect at the intermediate level than at the ground level. This is a condition called "population inversion," which is indicative of a state of great potential energy. This energy is released when a random chromium atom drops from the intermediate level to ground level, causing a photon to be emitted. This photon collides with nearby metastable atoms, causing the emission of more photons which in turn collide with other metastable atoms. Some of these photons strike the silvered ends of the ruby crystal and are reflected along the rod to the opposite end. In a short time, a whole stream of photons is bouncing back and forth between the ends of the rod. As they do so, they trigger additional metastable atoms and eventually generate an intense beam of light which, when of sufficient magnitude, bursts through the lightly silvered end of the rod as a pulse of monochromatic (one color or frequency), spatially coherent light. Because the photons in this beam of light are almost exactly parallel, they remain in this state for great distances, whereas incoherent beams of light, such as emitted from an incandescent lamp, are quickly dispersed in all directions. The basic operation of the ruby laser is similar to that of other types of lasers, even though the material that is irradiated differs.

A laser beam, with its coherent light waves, is capable of some astounding feats. It can, for example, be focused with enough intensity to heat a pail of water at a distance of one thousand miles. In 1962, laser beams projected toward the moon's surface arrived with a spread of only two miles after having traveled some 240,000 miles through space. In medicine, the laser can prevent the eventual blindness caused by a detached retina: a laser beam focused through the lens of the eye welds the retina back in place. Eye tumors can also be disintegrated by a laser beam, and lasers can be used to cauterize wounds and to take the place of the dentist's drill.

Communications

One possible application that has resulted in intensive research is the use of lasers in the transmission of information. The tremendous bandwidth available at coherent light frequencies is many times that presently available in the microwave spectrum. Utilization of the light spectrum for communica-

tions brings closer to the realm of reality the possibility that each person on earth can someday have his own personal wavelength.

To use light frequencies for carrier communications, it is necessary to modulate the carrier frequency. Optical modulators for communications are still under investigation. Optical waveguides capable of conveying light from point to point offer great possibilities for the transmission of information by laser. At present, a point-to-point laser communications system over open space similar to a microwave system seems unlikely, since light waves are much more subject to interference by fog, clouds, and dust than are microwaves. There is, however, a possibility that the carbon-dioxide laser, which produces continuous high power laser beams in the infrared region, may make atmospheric communications possible.

Semiconductor Lasers

Many lasers are inefficient in the conversion of pumping energy into laser light. Consequently, they tend to require great amounts of power for operation, which usually means a very heavy and unwieldly power supply. The semiconductor laser (also called injection, junction, and diode laser) is a step toward the solution of this problem, since it is extremely efficient and lightweight. The injection laser is simply a semiconductor diode, usually composed of gallium arsenide (GaAs), or of gallium arsenide-phosphide (GaAsP). The semiconductor gallium phosphide fascinated scientists for years because it emitted flashes of bright red light when current passed through it. It was later discovered that other semiconductor materials composed of compounds of gallium and arsenic radiated infrared light when a current was passed through them. And, while this infrared light consisted of incoherent waves, it still was possible to use it in an experimental transmission system to send messages through the atmosphere to a receiver 30 miles away. The fact that semiconductors were capable of converting electrical energy into photons of electromagnetic radiation led scientists to believe that with a large enough current, lasing action might be achieved. In 1962, laser action in a semiconductor was achieved, although a current density of 10,000 amperes per square centimeter was necessary for proper operation. This required that the laser be cooled down to temperatures close to absolute zero (–273°C).

An injection laser might appear as shown in Figure 5. Electrons injected into the diode junction produce the initial photons which go on to strike neighboring atoms, causing them to emit additional photons. Just as in the ruby laser, the end surfaces of the semiconductor junction are polished so that they may reflect the photon bombardment; when the reflections build up to sufficient energy, a beam of coherent light is emitted from the semiconductor junction. One definite advantage of the injection laser is that there is no requirement for a separate pump mechanism since the direct injection of electrons performs this function (for almost each electron injected, a corresponding photon is produced). Also, the output frequency of the injection laser can be controlled by varying the operating temperature and by altering the chemical composition of the semiconductor material. The output power varies in proportion with the magnitude of the current through the device.

In spite of the development of ruby, gas, and semiconductor lasers, and their subsequent use in industry, medicine, astronomy and as laboratory tools, the role of the laser as a signal carrier has as yet to be truly realized.

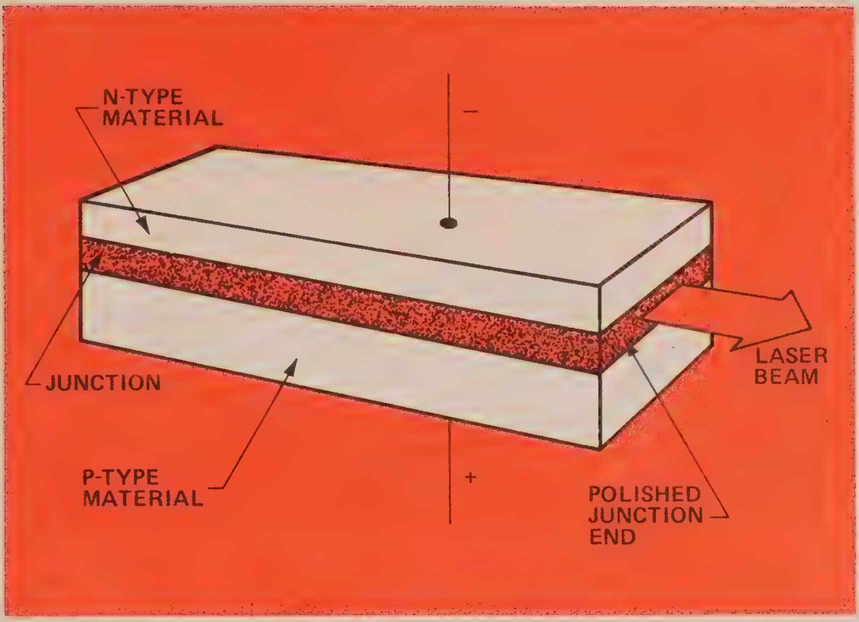


Figure 5. Principal elements of the injection laser.

The large lasers are too expensive for a large-scale communications system, and even the injection lasers have, until recently, had such a high current requirement that continuous operation could not be sustained (due to overheating) except at cryogenic (very low) temperatures. At room temperature, injection lasers could operate properly only in a pulsed mode, a limitation that is not compatible with the modern transmission of information.

Continuing developments in laser technology have produced semiconductor lasers capable of operating continuously at room temperature, and with a significantly smaller current requirement than the previous injection lasers. The types of lasers which can perform these functions are called double heterostructure (DH) injection lasers. In fact, the most promising injection laser for telecommunications is presently the double heterostructure laser composed of layerings of $Al_xGa_{1-x}As$ (with varying values of x). This type of laser is well suited for use with optical fiber communications systems. It has such desirable characteristics as small size, ruggedness, good efficiency, and can be pumped and modulated by means of the injected current. DH lasers are tinier than a grain of sand, can be powered by ordinary dry cell batteries, and together with other necessary advancements in optoelectronics will very likely play a part in the first experimental optical transmission systems.

The coupling of laser energy into multimode optical fibers can be accomplished easily and efficiently simply by positioning the fiber at the laser output end. However, there are still problems to be solved. For example, when coupling a DH laser to single-mode optical fiber, some type of mode transformation device is necessary because the mode of the laser output has an elliptical cross section, while that of the optical fiber is circular.

The future of laser communications is still very much in the making. And, at this point, it is difficult to say just where, and how far, scientific development will take the field of optical communications. But in a world where modern technological miracles seem to be occurring in great abundance, it is perhaps unwise to set limitations on possibilities by calling them science fiction.

BIBLIOGRAPHY

1. Asimov, Isaac. *Asimov's Guide To Science.* New York: Basic Books, Inc., 1972.

2. Brotherton, M. *Masers and Lasers.* New York: McGraw-Hill Book Company, 1964.

3. Brown, Ronald. *Lasers: Tools of Modern Technology.* New York: Doubleday & Company, Inc., 1968.

4. Carroll, John M. *The Story of the Laser.* New York: E.P. Dutton & Co., Inc., 1970.

5. Fulenwider, John E. "Some Potential Impacts of Optical Fiber Transmission and Integrated Optics on World Communications," *1975 IEEE Intercon Conference Record,* Session 5, April 8 – 10, 1975.

6. Miller, Steward E., Tingye Li, and Enrique A.J. Marcatili. "Research Toward Optical-Fiber Transmission Systems," Part II, *Proceedings of the IEEE,* Vol. 61, No. 12 (December 1973) 1726 – 1730.

7. Panish, Morton B. "Heterostructure Injection Lasers," *Bell Laboratories Record,* November 1971.

8. Panish, Morton B., and Izuo Hayashi. "A New Class of Diode Lasers," *Scientific American,* July 1971.

9. Schawlow, Arthur L., ed. *Lasers and Light: Readings From Scientific American.* San Francisco: W.H. Freeman and Company, 1969.

10. Smith, V. William, and Peter P. Sorokin. *The Laser.* New York: McGraw-Hill Book Company, 1966.

GTE LENKURT
DEMODULATOR
SEPTEMBER/OCTOBER 1975

BASIC SEMICONDUCTOR MEMORIES

PARTS 1 AND 2

BASIC SEMICONDUCTOR MEMORIES Part 1

Virtually all data processing equipment is concerned with the storage and transfer of information in a digital form. The essential characteristic of the elements comprising such equipment is operation in at least two well-defined and distinct states.

Data processing systems have revolutionized our world, allowing vast amounts of information to be stored, exchanged, updated, and utilized in ways undreamed of a few years ago. A chain of department stores can be tied together, for example, by a data network making current inventory, personnel and credit information instantly available at widely separated locations; a railroad can control activity at its switchyard from a control center hundreds of miles distant; switching functions within a telephone company office can be accomplished rapidly and reliably in accordance with stored program instructions. The information involved in each of these examples is different, but the underlying processing principles are the same.

A digital computer or data processor of any type is basically a stored-program machine, in which a memory facility holds a set of operating instructions – the system program. Information is put into a digital format (see the August, 1974, *Demodulator* for a discussion of this conversion) and fed into the machine, which retains it in a data memory. The instructions in the program memory, which are also in digital form, tell the processor what problem is to be solved, or function performed, related to the input data. When the operations demanded by the program memory have been completed, the data is fed out to be utilized in some manner. In practice, program instructions are often stored within the same memory facility as the data; in this way, the program can also be changed if desired.

Despite surface differences, the ways in which memory facilities receive, hold, and feed out information are all based on either sequential or random access principles.

Sequential Memories

The sequential, or serial, memory method requires that the data bits comprising the information be arranged in a particular order. Data stored in this manner – including both programmed instructions and input information – is retrieved strictly in accordance with its position in a time sequence.

A simple example of sequential memory storage is paper tape, which uses the presence or absence of holes to indicate the condition of a data bit; the combined states of several bits identify a particular digit (see Figure 1). The tape is moved through a sensing device to convert the hole patterns into a series of pulses for electronic processing. Each piece of information is retrieved as it passes the

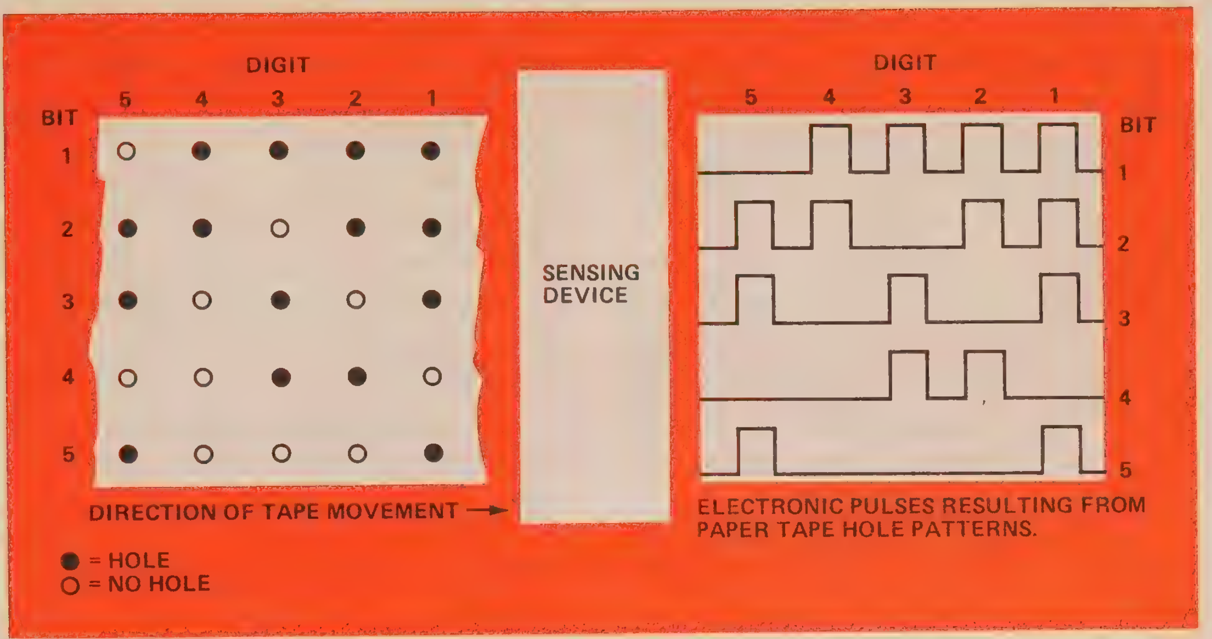


Figure 1. Paper tape is a form of sequential memory storage, in that the bits identifying a given digit can only be retrieved during their allotted sensing time, which may cause delay in finding desired data.

sensor. Another form of sequential memory storage is magnetic tape, which stores information as magnetic flux variations corresponding to the 1's and 0's of digital data.

Sequential access memory systems, including punched cards and magnetic disks as well as tapes, have been widely used in the area of mass computer memories. They typically contain program instructions and must be physically introduced into the processor system — threaded through a sensing device, for example — so that input data can be operated upon. They may also be used to retain the data for future processing. These devices provide a permanent storage capability since the cards, disks, and tapes can be removed and filed for repeated use, and they have non-destructive readouts (i.e., data does not have to be re-entered every time it is used); however, they have some shortcomings which limit their usefulness in high-speed memory applications.

For example, sequential access can be a relatively slow process because a large amount of irrelavent data may have to be scanned before the desired bits are found. This delay may be from milliseconds to minutes, which is much too great for many of the uses to which modern data processing equipment is put. Additionally, these sequential mass memory systems require a mechanism to move the data-carrying medium past the sensor; this device is of necessity completely mechanical, and is subject to the adjustment and maintenance considerations which apply to all such devices.

Temporary Memories

Cards, tapes and disks continue to play an important role as high-density, long-term mass memory storage elements in such applications as personnel record maintenance, inventory control, and retention of performance data for comparison with future achievements.

For low capacity, temporary memory storage, however, structures composed of semiconductor devices have become dominant in recent years.

Temporary data storage facilities are used in such areas as office equipment (calculators, etc.) and immediate-use or "working" memories wherein a data processing device can hold the data with which it is dealing at any given time. It is not uncommon for a data processing system to use tapes and similar elements as permanent, high-volume program storage facilities, and semiconductor structures for the "working" memories in which the stored data is operated upon.

Temporary memories are also widely used in data processing terminals, which serve as remote input/output units for a large central computer and may be in any number of forms, from a simple typewriter keyboard to a small computer. Terminals provide a means of encoding data for manipulation by a central computer, and decoding it for use by a human operator.

Interconnection of central computer and remote terminal is commonly made over telephone lines through an interface unit called a modem, or data set. The data set may also contain a temporary memory facility which allows it to hold data and either condition it for transmission over the lines or prepare received data for application to the processor.

The semiconductor devices used in temporary memory structures are typically arranged as groups of individual units called memory cells, each of which stores one bit of information as either a logic 1 or logic 0. A cell may consist of as little as one transistor-capacitor combination, or it may be a complex arrangement of several components, but, whatever its composition, it has at least two states that can represent digital data bits.

Shift Registers

A shift register is a device for the temporary storage of digital information; when a shift, or clock, pulse is applied, the register accepts new data and moves every stored bit one step toward the output.

Figure 2 shows an example of a semiconductor shift register containing five memory cells, each of which is a solid-state reset-set (R-S) flip-flop circuit. Data applied to the register input in a digital form is stored and shifted from cell to cell in accordance with the clock cycle rate. A logic 1 on the S input of cell 1, for example, will set the flip-flop to the 1, or "on," state if a clock pulse is present at input C, thus storing the digit. On the next clock pulse, the stored bit is shifted out of cell 1 to set the second cell to the 1 state, and a new digit is fixed in the first cell. This procedure continues through the register, with the output of the final cell being shifted out for data processing. The clock frequency controls the rate at which the shift register stores and feeds out data bits, with each bit being delayed between input and output by as many clock cycles as there are cells in the register. Since the storage and retrieval are done on a first-in, first-out basis, the shift register is a sequential memory storage device.

Digital data can also be handled by a random access process. This does not mean, of course, that no orderly procedure is involved; it means, rather, that information can be stored in a particular memory cell, or location, and retrieved without regard for any other location.

Random Access Memory Systems

A random access memory (RAM) can be defined as a structure in which any data bit can be stored (written) or retrieved (read out) in any order.

One of the most basic ways to create a memory cell is through the use of the bistable multivibrator, or

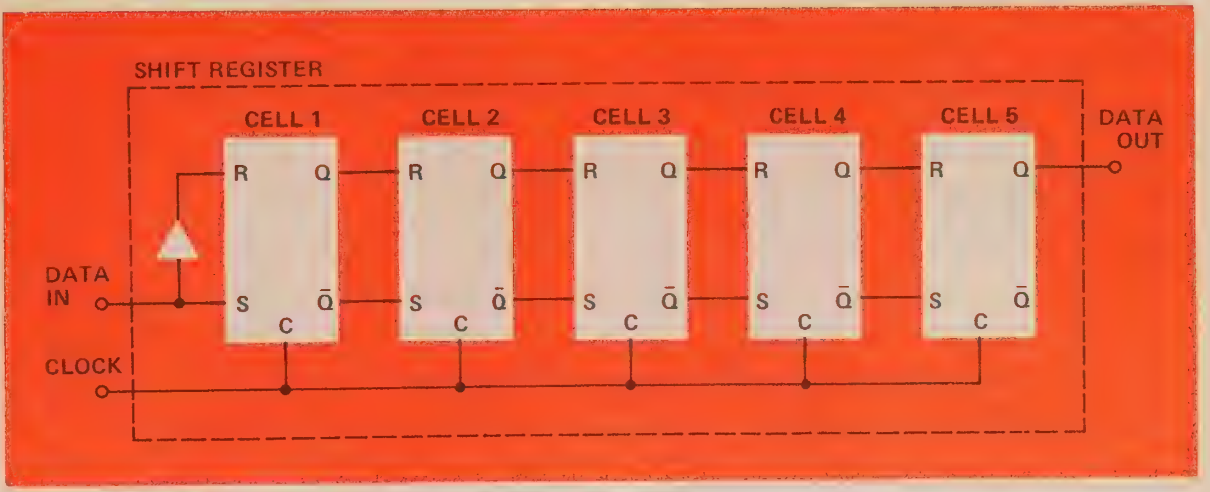


Figure 2. The shift register, which stores data for as long as it takes for the clock to move it through each cell, is typical of semiconductor sequential memory systems.

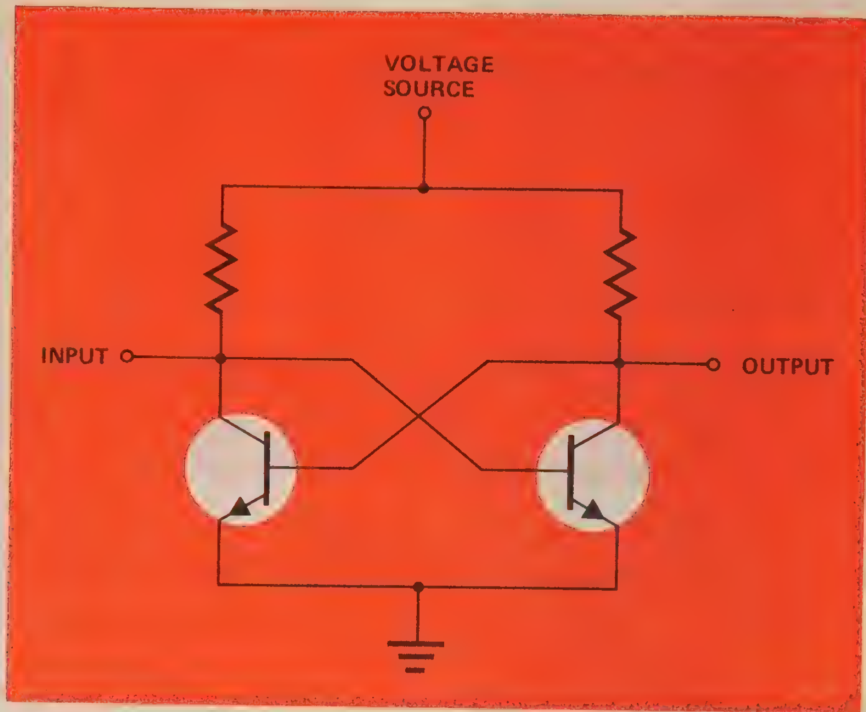
flip-flop. As shown in Figure 3, such a memory cell may consist of only two transistors, two resistors, and a power source. In this cell, one or the other of the transistors is always conducting, holding the other one off. When an external signal forces the off transistor into conduction, the initially on transistor turns off and remains in this condition until another external signal resets it. The flip-flop, therefore, has two stable states which can be used to store information in the form of logic 1's and 0's.

A RAM is essentially a matrix of such memory cells, with each cell identified by a unique code, or address. The data processing equipment can retrieve a bit of information by addressing the proper location. Because of the matrix structure, the time required to locate any given bit is approximately the same as that required to locate any other bit. For example, in Figure 4 the digit stored in cell 1, at location A1, could be available at the data output in almost exactly the same time as the bit in cell 16, location D4. This rapid access to information makes the RAM ideal for application as a temporary storage facility.

Random access memory structures are of two basic types: read/write and read only. A read/write RAM is programmable; that is, data can be entered into, changed, and removed from the memory at any time. A read only memory (ROM), however, has certain data patterns fixed into it, usually during the manufacturing process. In such a structure, information can be read out – it will always perform the same function – but the stored program does not change. Because the data pattern is fixed, a ROM retains its program regardless of circuit power considerations; that is, it is a "nonvolatile" memory device. A read/write memory, however, needs a constant source of power to remain in operation; if power is removed, the semiconductors stop conducting and the stored information is lost, so the read/write RAM is considered to be a "volatile" device.

Although it is not technically accurate to do so, common usage has led read/write memories to be referred to simply as RAM's, while read only structures – which are, in reality, a type of RAM – are designated ROM's, and this discussion will follow the same nomenclature.

Figure 3. The bistable multivibrator, or flip-flop, has two stable states, making it an ideal digital data memory storage cell.



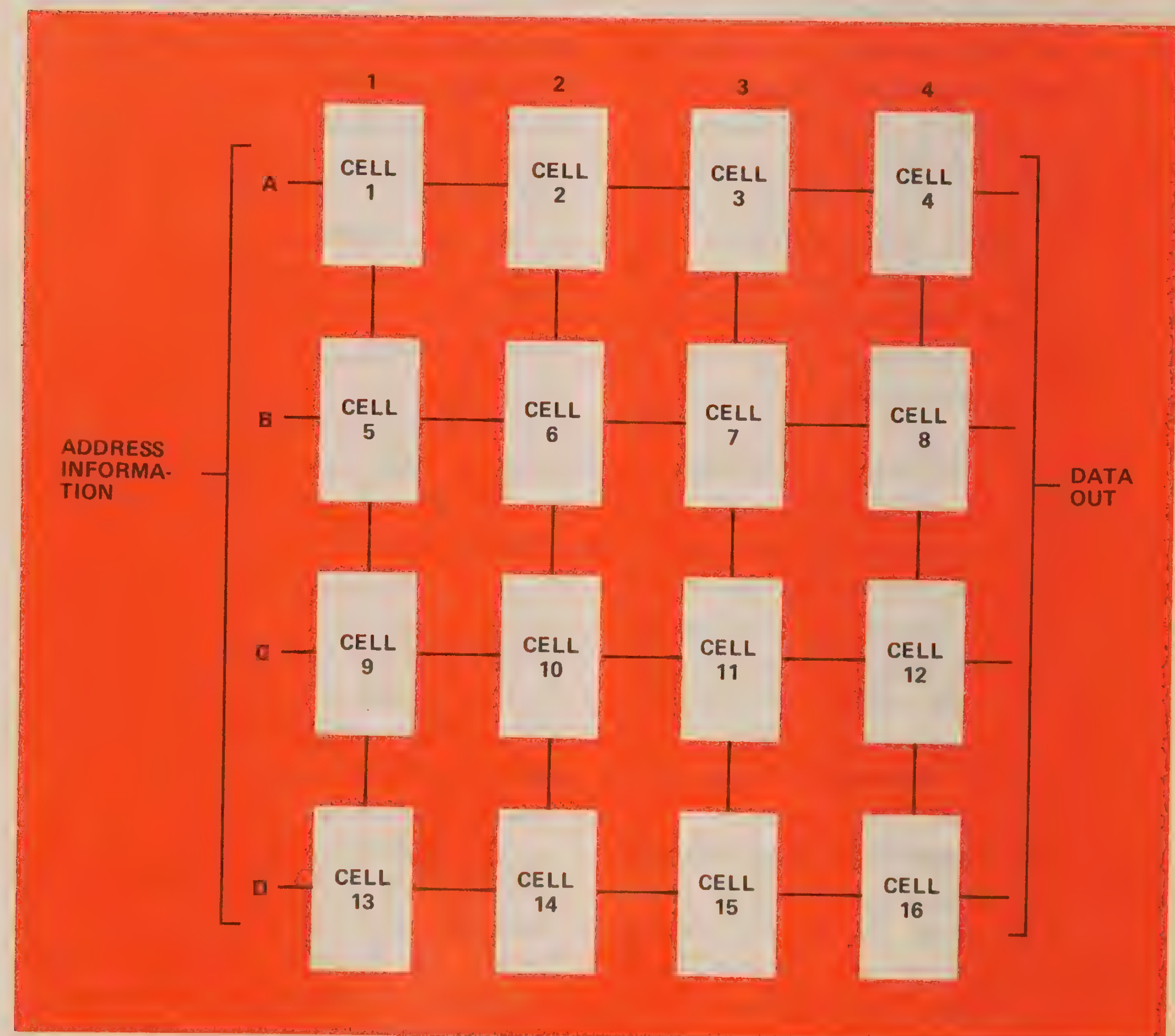


Figure 4. A random access memory (RAM) is a matrix of memory cells, any of which can be accessed without regard for any other cell.

Bipolar and Unipolar Transistors

The two most widely used devices in memory matrix design today are the bipolar and unipolar, or field effect, transistor. Each can be easily realized as an integrated circuit (IC) component, and they are readily adaptable to virtually any circuit configuration.

Essentially, a bipolar transistor is a semiconductor device whose conductive properties depend upon both majority and minority carriers; that is, current flows in a bipolar transistor because of the simultaneous movement of both positive and negative charges. Negative charges predominate in N-type semiconductor material because there is a surplus of free electrons within the material's atomic structure; P-type material, however, has a shortage of free electrons. The regions in which the electrons would normally exist act as positive, mass-bearing charges called "holes"; P-type material thus maintains an excess of positive charge carriers. The common transistor is a general type of bipolar device, since its current flows due to hole and electron movement. Figure 5A shows the normal flow pattern within an NPN structure (Figure 5B shows a bipolar NPN structure as an integrated circuit). The forward-biased emitter-base junction allows electrons to be injected by the emitter into the base region. Within the base, the greater part of the current flow is caused by holes combining with the excess electrons. The reverse bias of the collector-base junction allows electrons to pass into the collector region; because there are two N-type regions, electrons are the majority carriers, although the simultaneous action of the minority carrier holes is indispensable.

The field effect transistor (FET) is a unipolar device, in that its current flow is the result of the movement of only one type of carrier. In what is called the p-channel FET, holes are the majority carriers, while the carriers in an n-channel FET are electrons.

Figure 5C shows a p-channel FET structure operating in the enhancement mode, which is the most common operating mode for FET's. In this mode, there is no conduction within the device when the gate voltage is zero; the other mode of operation is called depletion, wherein the semiconductor device is always conducting and requires a proper gate-to-source voltage to turn off.

When the gate in Figure 5C is made negative with respect to the source, it creates an electrostatic field which attracts holes from the n-type semiconductor material toward the area directly below the gate dielectric material. Initially, this n-type area has a surplus of electrons, but as the holes are drawn into it, the electrons are neutralized. At some gate voltage, the holes become dominant and a current-carrying channel is produced between source and drain in which holes are the majority carriers. Because the gate is electrically isolated from the rest of the structure by the dielectric material, there is no current flow into it; the channel, which allows current to flow between source and drain, is created and maintained by the electrostatic field.

The need to provide more electronic function in increasingly small areas has resulted in a great variety of miniaturized circuits. Bipolar transistors, for example, can be realized as discrete items, such as those seen in various entertainment products. In data processing applications, however, the large number of components required to produce a memory matrix makes the use of such bulky devices impractical; a circuit board with several hundred discrete transistors mounted on it — which is what a memory

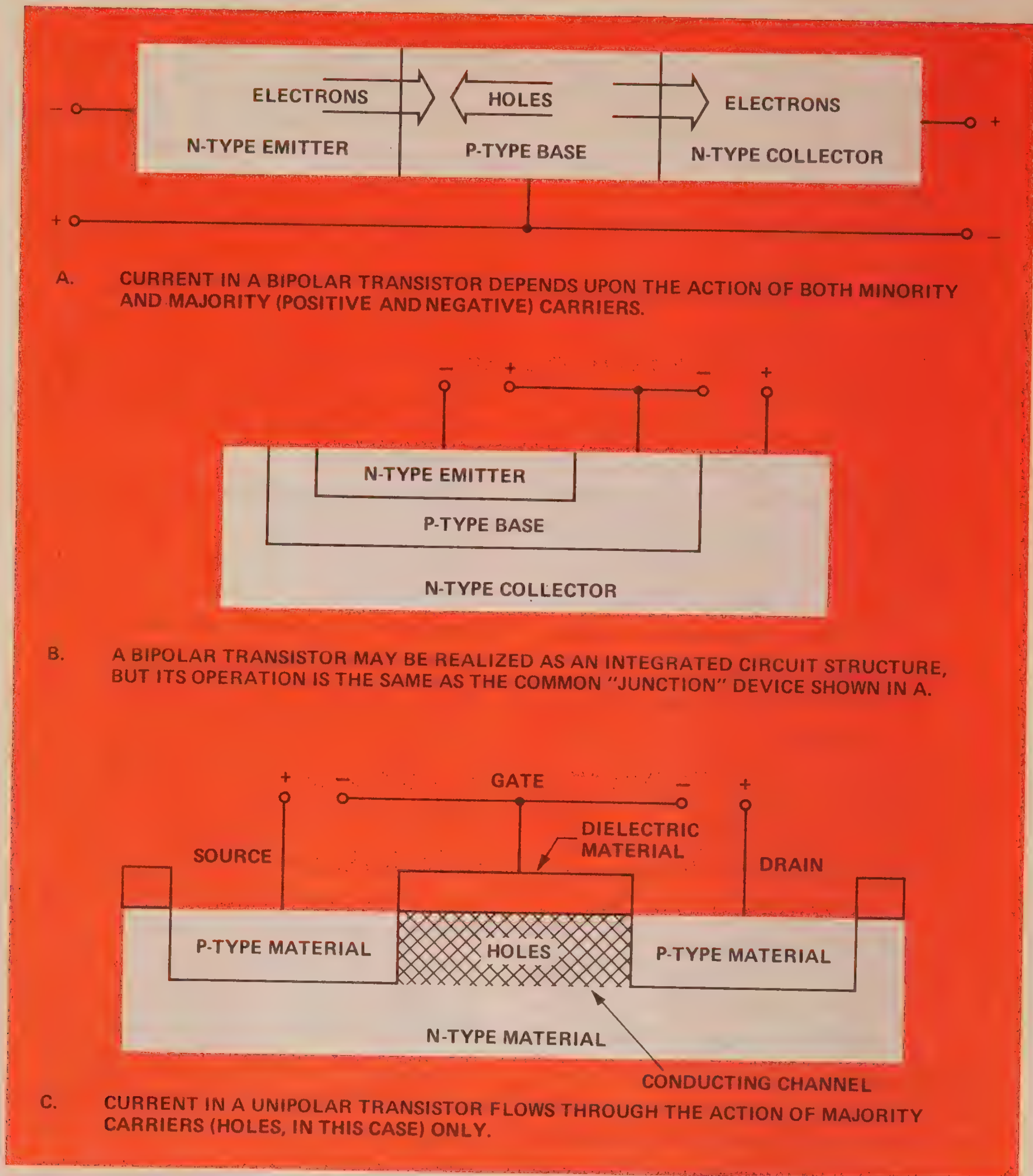


Figure 5. Bipolar and unipolar transistor structures are the most widely used semiconductor memory cell components.

matrix would require — would be too unwieldy to be of any real use.

Bipolar Transistor RAM's

Integrated circuit techniques allow quantities of circuit elements to be realized in a small space; these techniques have been used to produce bipolar transistor memories of various microminiaturized sizes and densities.

The basic storage element in these matrices is the bistable flip-flop, which appears in many circuit variations to meet different application requirements.

A bipolar RAM cell which has been widely used is the transistor-transistor-logic (TTL) type, exemplified by the multiple-emitter circuit (see Figure 6). In this case, the data processor applies

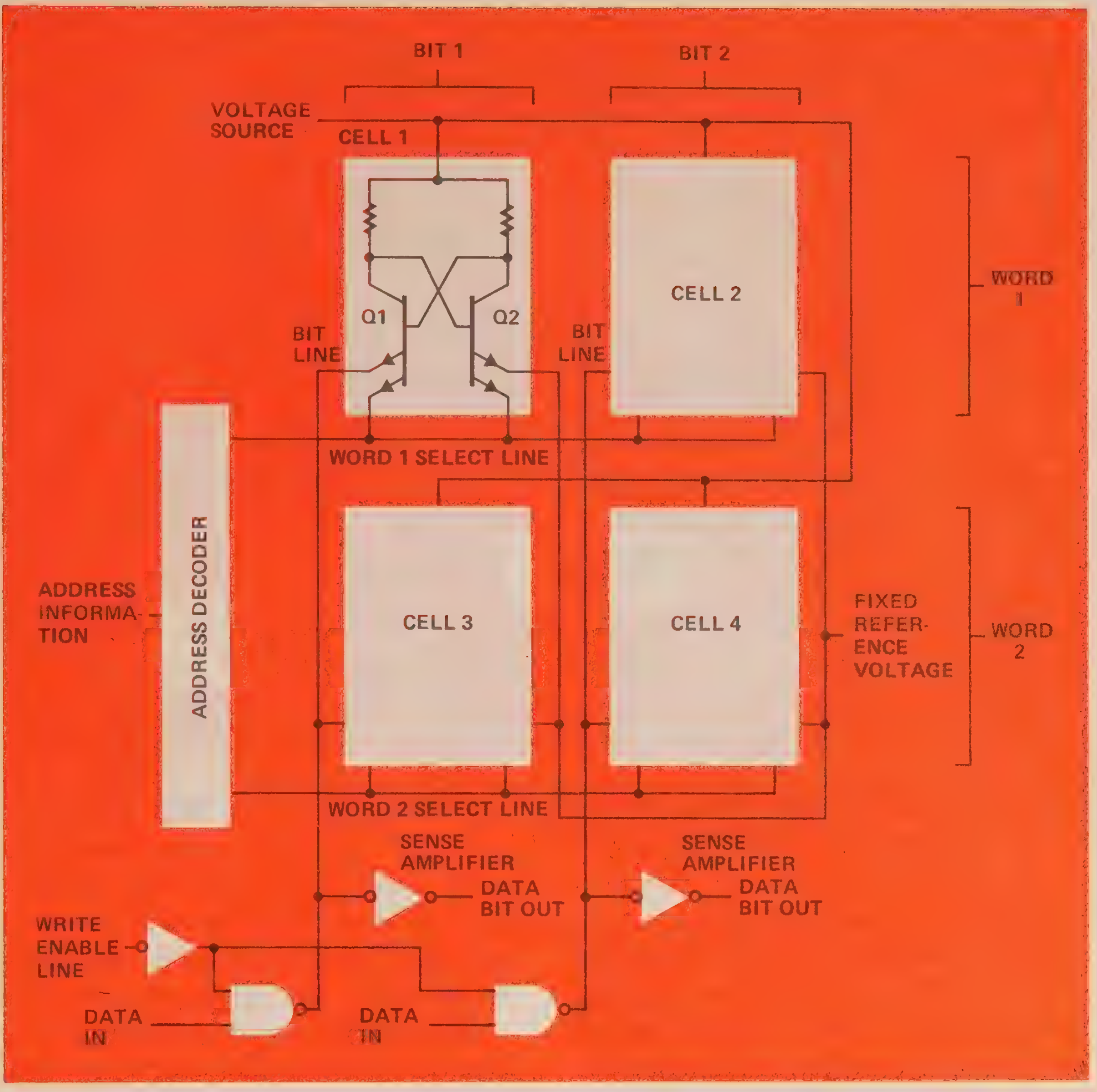


Figure 6. A 2-word, 2-bit-per-word memory array utilizing multi-emitter bipolar transistor structures.

an address code to a decoding circuit; the decoded address raises the voltage on the correct word select line (places it at a logic 1 level), preparing the cell for the read or write function. A logic 1 bit can be written into the cell, for example, by placing the write enable line at a low voltage (logic 0) level while the data bit input is logic 1. This causes the bit line to be low, turning Q1 on and Q2 off, a state which represents a logic 1 within the cell. Once the write function is complete, the address changes, the word select line returns to a logic 0 state, and Q1 remains on. To read out the stored digit, the address raises the word select line level and the write enable line is held high (logic 1), allowing read current representing the value of the cell's contents to flow into the appropriate sense amplifier for output to the data processor. Because the read process does not change the state of the flip-flop, the stored information is not lost and the cell is considered to have a "non-destructive" readout capability.

The 2-word, 2-bits-per-word memory shown in Figure 6 is, of course, limited in its application. The same

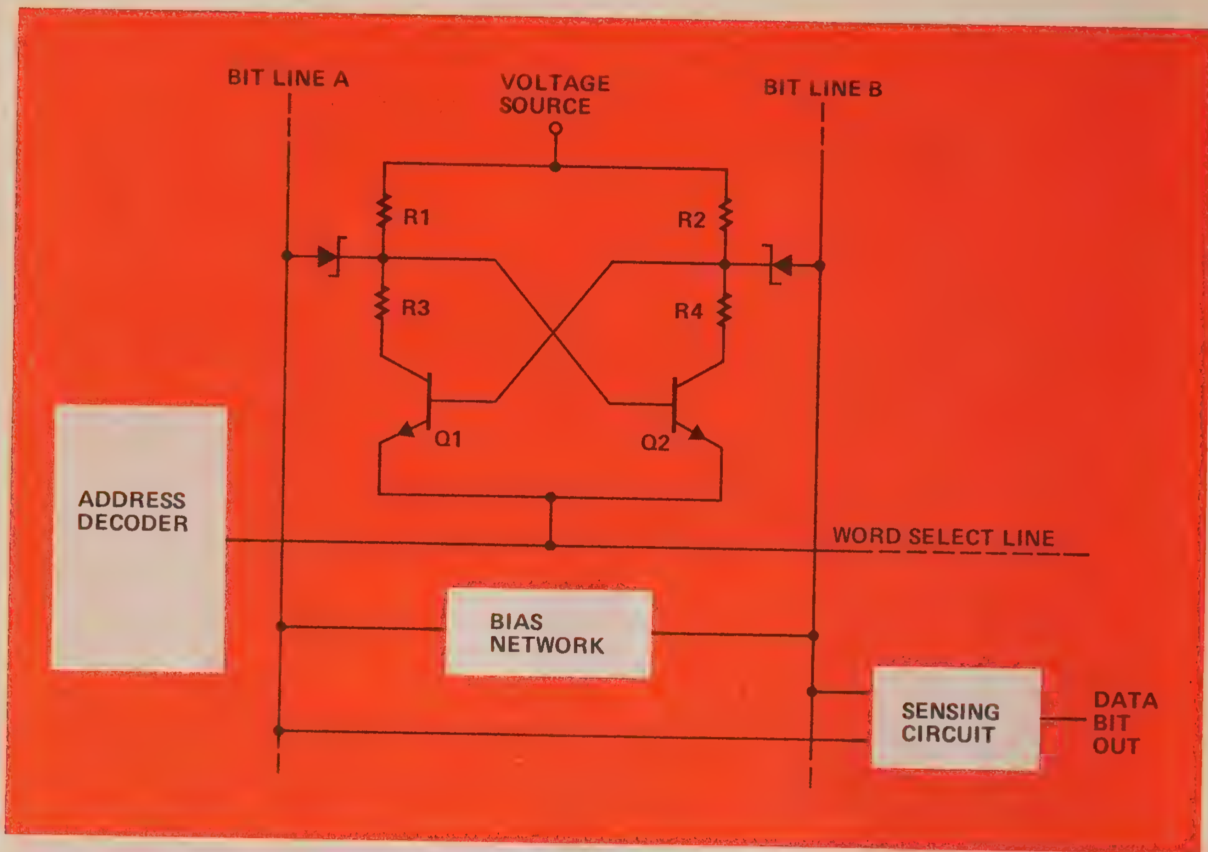


Figure 7. A diode coupled memory cell uses gating diodes to control conduction and reduce power consumption.

addressing, reading, and writing functions, however, are performed in bipolar RAM's containing several times the number of cells. A typical example of expanded capacity is a single integrated circuit capable of storing 16 words of 4 bits each, for a total of 64 bits on one tiny silicon chip. GTE Lenkurt uses nine such chips in both its 262A and 262B data sets. The bipolar RAM's comprise a data memory, in which input data is held to be operated upon. Because of the read/write capabilities of the RAM's, the data being processed can constantly be updated and changed.

Major considerations in memory design include the speed with which a cell can be made to change state (access time) and the amount of power dissipated by the cell's components.

Operating in a saturation mode – in which the "on" transistor constantly conducts the maximum possible current – the TTL-type memory cell requires some amount of time to drive the transistor out of saturation before a change of state can occur. This delay is only on the order of nanoseconds, but is enough to concern circuit designers. In addition, the saturation mode consumes relatively large amounts of power. Two of the more successful configurations developed to overcome these disadvantages are the diode coupled and emitter-coupled logic (ECL) cells.

Diode Coupled RAM's

Two gating diodes are used to control conduction in a diode coupled cell (Figure 7). In integrated circuits, these diodes are frequently "hot-electron," or "Schottky barrier," devices, which become forward-biased at lower voltages than conventional diodes.

If the state of a cell must be changed to store a bit, the address

decoder causes the voltage on the word select line to be forced low, while the voltage is raised on the bit line associated with the transistor to be turned off. Referring to Figure 7, in which Q2 is hypothetically to be turned off, raising the bit line B voltage and dropping the word select line (effectively making it more negative) draws additional current through R4, increasing the base voltage on Q1 to the point at which it begins conducting. The cross-coupling of the transistors then causes Q2 to turn off, thus effecting the cell's change of state.

Reading the stored digit out of a diode coupled cell also requires that the word select line be forced low, but in this case there is no voltage increase on either of the bit lines. The combined effects of the lowered word select line and a bias network cause the diode associated with the "on" transistor to be forward-biased, causing the diode to conduct. Since the diode associated with the "off" transistor is reverse biased, a differential voltage develops between the bit lines. A sensing circuit determines the cell's logic state from this voltage.

Read current in a diode coupled cell is greater than standby current, which flows when the cell is storing a bit without being addressed, but is substantially lower than write current. Because of this, the voltage developed across the load resistors during the read operation is not great enough to change the cell's state and the readout is a nondestructive process.

Since standby current is lower than read current and is present a greater percent of the time, overall power consumption in a diode coupled cell is lower than that of a TTL device.

ECL RAM's

The structure of emitter-coupled logic (ECL) memory cells closely resembles that of TTL cells, but biasing techniques are used to keep the transistors out of saturation. This allows the ECL storage element to change state very rapidly; the greatest advantage of ECL over other semiconductor memory configurations is that it has the shortest access time of all. Reading and writing processes are accomplished in essentially the same manner as for TTL, but at a greater speed. Because it constantly draws high current, however, an ECL memory cell has even higher power consumption than TTL, a fact which does impair its usefulness in certain applications.

Numerous other semiconductor memory cell configurations utilizing bipolar transistor integrated circuit technology have appeared. Virtually all of these are variations on the three basic types described here. The October, 1975, *Demodulator* will continue this discussion, describing some of the basic metal oxide-silicon (MOS) memory structures.

BIBLIOGRAPHY

1. Altman, L. "Semiconductor Random-Access Memories," *Electronics*, Vol. 47, No. 12 (June 13, 1974), 108-110.

2. Eimbinder, J., ed. *Semiconductor Memories*. New York: Wiley-Interscience, 1971.

3. Luecke, G., *et. al. Semiconductor Memory Design and Application*. New York: McGraw-Hill Book Co., 1973.

4. *Reference Data For Radio Engineers*, Howard W. Sams & Co., 1968.

BASIC SEMICONDUCTOR MEMORIES
Part 2

Almost every new computer utilizes semiconductor memory devices, and the use of these devices is spreading to the communications field. Semiconductor memories are beginning to replace cores and drums in telephone switching centers, and digital data carrier systems using PCM techniques, which are inherently suited to semiconductor logic and memory circuits, are being readied for introduction.

The September, 1975, *Demodulator* introduced many of the concepts basic to the understanding of semiconductor memories, and described various types of bipolar transistor read/write random access memory (RAM) cells. This issue concentrates on the discussion of metal oxide-silicon devices.

MOS Technology

One of the major objectives in semiconductor memory design has been to incorporate as much capacity as possible in the smallest area. The greatest size reductions have been achieved with metal oxide-silicon (MOS) techniques, which produce field effect transistor (FET) structures that are considerably more compact than the bipolar integrated circuits (IC's) previously discussed.

An MOS FET is formed by depositing an insulating metal oxide – most often silicon dioxide – on a chip of silicon. Etching processes then remove the oxide from selected areas of the chip, exposing the substrate at source and drain locations while leaving the gate region insulated. Further processing establishes n- and p-type areas within the substrate.

The size reduction possible with MOS techniques allows a much denser memory array to be produced within a given space than is possible with bipolar devices; there are also substantially lower power requirements and reduced packaging costs.

Storage cells composed of MOS FET's may be of either a static or dynamic nature. A static cell retains its stored data as long as power is supplied to the circuit; a dynamic cell depends upon capacitive charge storage to hold its data, and must receive a "refresh" input to counteract the effects of leakage.

Static MOS RAM's

The basic static MOS RAM cell is a bistable multivibrator (see Figure 1) closely resembling the bipolar flip-flop used in TTL memories. In an MOS flip-flop, however, transistors serve not only as cross-coupled inverters (Q3 and Q4), but also as load resistances (Q1 and Q2). Electrical isolation of the FET gate results in a very high input resistance which can be controlled by the gate voltage. A large-value resistor can thus be produced by an FET in a relatively small space compared to a conventional resistor.

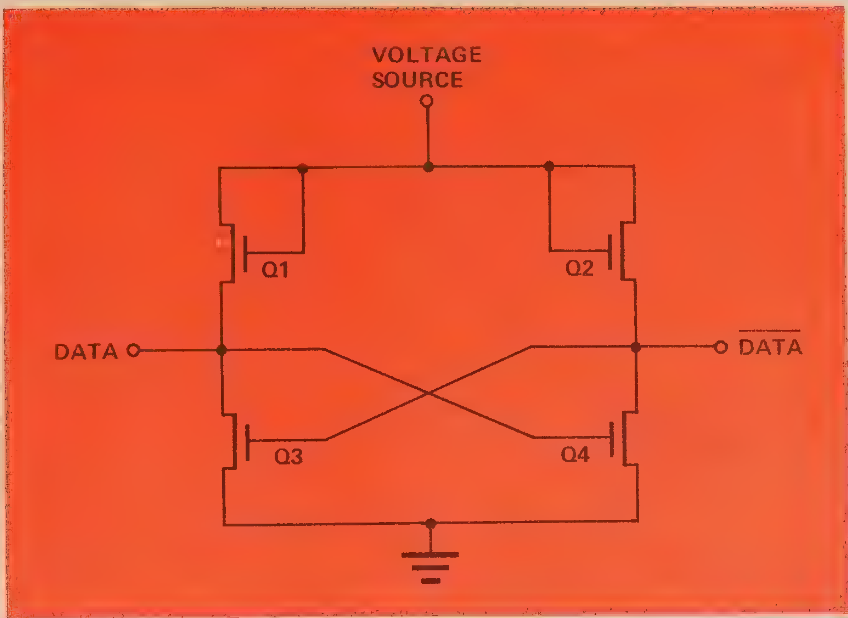


Figure 1. The heart of the static MOS RAM storage cell is the bistable flip-flop composed entirely of field effect transistors (FET's). The logic level at one terminal is always the complement of that at the other.

Since only one of the cross-coupled inverters conducts at any given time, the cell has two stable states which can be used to store information in the form of logic 1's and 0's. The state of the cell is determined by external address and data signals. The cell's state remains constant unless changed by an external signal, so no refresh action is required and the circuitry needed to support the operation of the cell is simplified.

A static MOS RAM storage unit, however, contains a minimum of four transistors, so it occupies a considerable amount of space on a silicon chip and consumes a relatively large amount of power. Because of these disadvantages, static devices have been largely replaced by dynamic MOS RAM's.

Dynamic MOS RAM's

The basic storage element in a dynamic MOS RAM cell is a capacitor, which holds and releases a stored charge in response to read and write commands. While the capacitor could be an external device, it is much more common for dynamic RAM's to utilize the capacitance existing between gate and source of the MOS FET itself. This capacitance is due to the isolation of the gate from the rest of the structure by a dielectric material. Charging the gate-source capacitance sufficiently to turn the transistor on represents a logic 1 state in most applications, while a lower charge or no charge at all serves as a logic 0.

Inevitably, as with all capacitive devices, the charge stored in the gate-source region drains off due to leakage current. If the charge is allowed to deteriorate too much, the data bit is lost, so some means must be provided to periodically restore, or refresh, the charge; a common requirement is that every cell in a memory matrix be refreshed every 2 milliseconds. Circuits to accomplish this are included in dynamic RAM designs, as are address decoding circuits.

The operation of a typical dynamic MOS RAM cell can be illustrated with the 3-transistor cell shown in Figure 2. In this case, information is stored as a charge in the gate-source capacitance (C_G) of transistor Q2. To write a data bit into this cell, an address decoder produces a write select signal, activating transistor Q1 and allowing data on the write data line to be transferred to the storage element. Depending

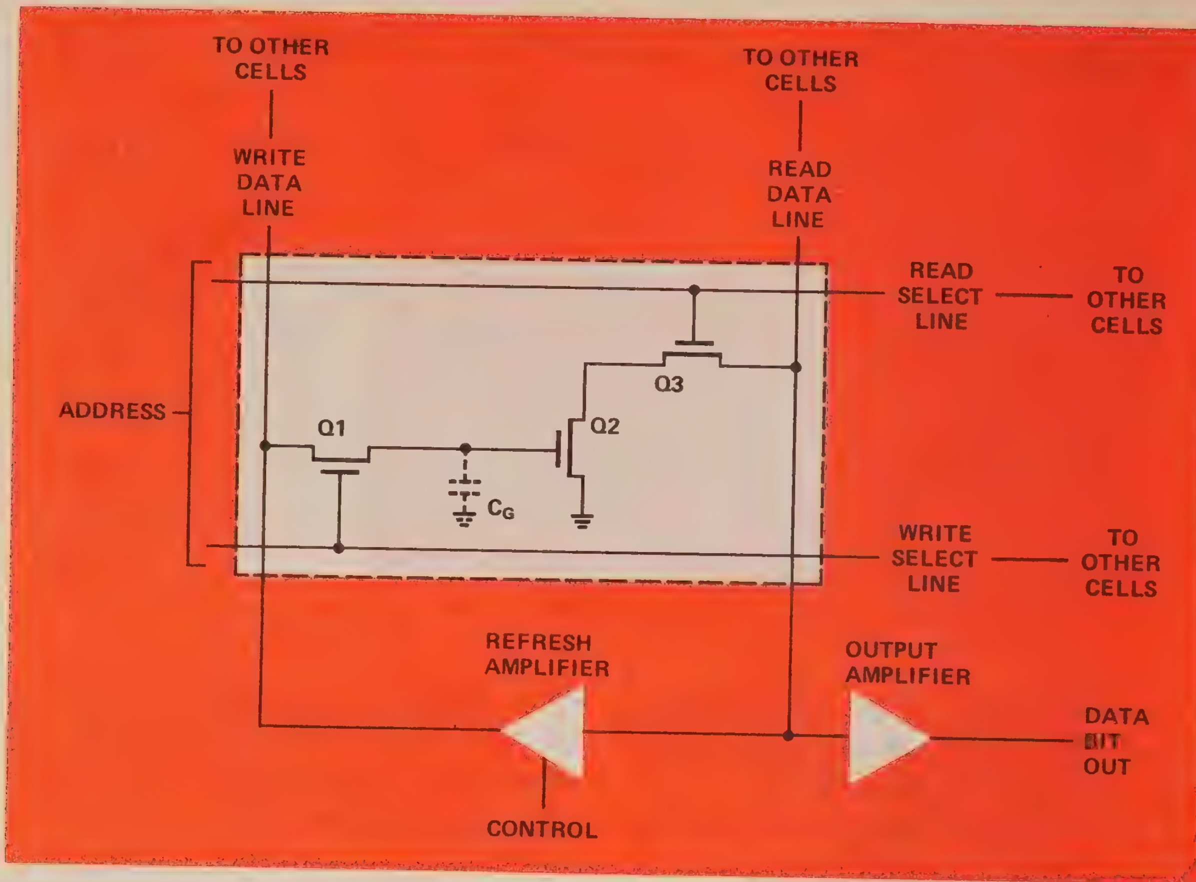


Figure 2. A dynamic MOS RAM cell stores data in the gate-source capacitance of one of its transistors.

upon the state of the data input, C_G either charges or is discharged. When the write select signal is removed at the end of the write cycle, the bit is held in the cell.

At the beginning of a read cycle, both the read and write data lines are preset to some voltage. When the address decoder produces a read select signal, Q3 is ready to begin conducting. If the charge on C_G is sufficient (logic 1), Q2 turns on and current flows through Q2 and Q3, reducing the voltage on the read data line. With no charge on the capacitance, Q2 and Q3 remain off and the read data line stays at its preset level. Because of the gate isolation, C_G is in the same condition (charged or discharged) at the end of the read cycle as at the beginning, making the read process "nondestructive." An output amplifier senses the state of the read data line and determines what cell condition would produce it (a low-level line generally indicates a stored logic 1) for application to the data processor.

Refresh

Refresh of the stored digit in Figure 2 is accomplished through a clocked amplifier connected between the read and write data lines. Control circuitry provides the timing necessary to keep the refresh cycle separate from the read and write operations.

The refresh process involves reading out the stored digit and writing it back into the cell. To do this, both data lines are preset at the beginning of the refresh cycle. A read select signal is then produced, transferring the bit to the read data line in the same manner as the normal read operation. The refresh amplifier inverts the condition of the read data line and applies it to

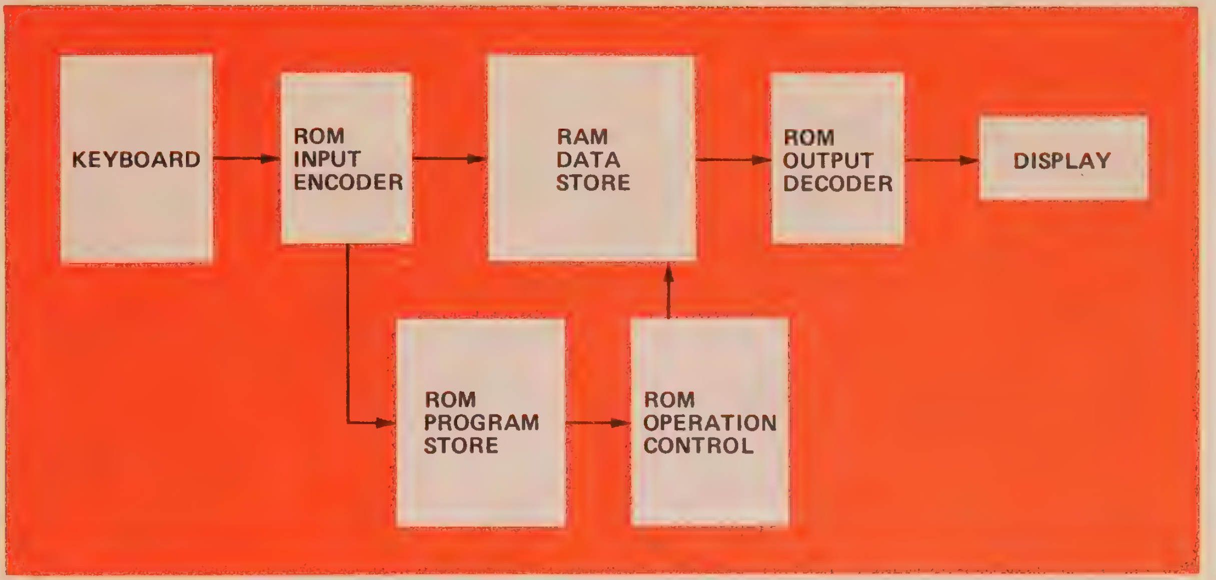


Figure 3. A pocket calculator typically utilizes both RAM and ROM facilities to process input data.

the write data line. A write select signal then replaces the read signal and the data present on the write data line is entered into the memory. If, for example, a logic 1 (maximum charge) is stored on C_G, the read data line is forced low (logic 0) when the read select signal forces Q2 and Q3 into conduction. The refresh amplifier inverts this and applies logic 1 to the write data line; the presence of the write select signal causes this data to be written into the cell as a refreshed bit. With no charge (logic 0) on C_G, this sequence is repeated, with a logic 0 appearing on the write data line to ensure that the capacitance is not charged by stray circuit currents.

In Figure 2, timing from the control circuitry allows a single amplifier to serve an entire column of cells. One alternative configuration uses a common read/write data line. This lets the cell form a loop within itself and thus eliminates refresh amplifiers.

ROM's

A read only memory (ROM) is a data storage facility into which information is normally written only once. After this entry, a ROM always produces the same output when addressed.

The difference between the read/write RAM and the ROM can perhaps be best illustrated with the example of the pocket calculator. In almost all calculator designs, a RAM matrix serves as a "working," or data, memory and ROM's are used for input/output interface, timing control and program storage (see Figure 3).

Each key on the calculator keyboard is identified by a unique binary number; all of these numbers are permanently fixed in the ROM encoder so that, when a key is pressed, the corresponding binary number appears as the encoder output. If a digit key is pressed, the bits comprising the number are written into the RAM data store. Function key (addition, subtraction, etc.) numbers are applied to the ROM program store as addresses. In the program store are contained instructions for each function; when an address is presented, the proper instructions are read out of the ROM, leading to performance of the desired operation upon the data held in the RAM. When the function has been

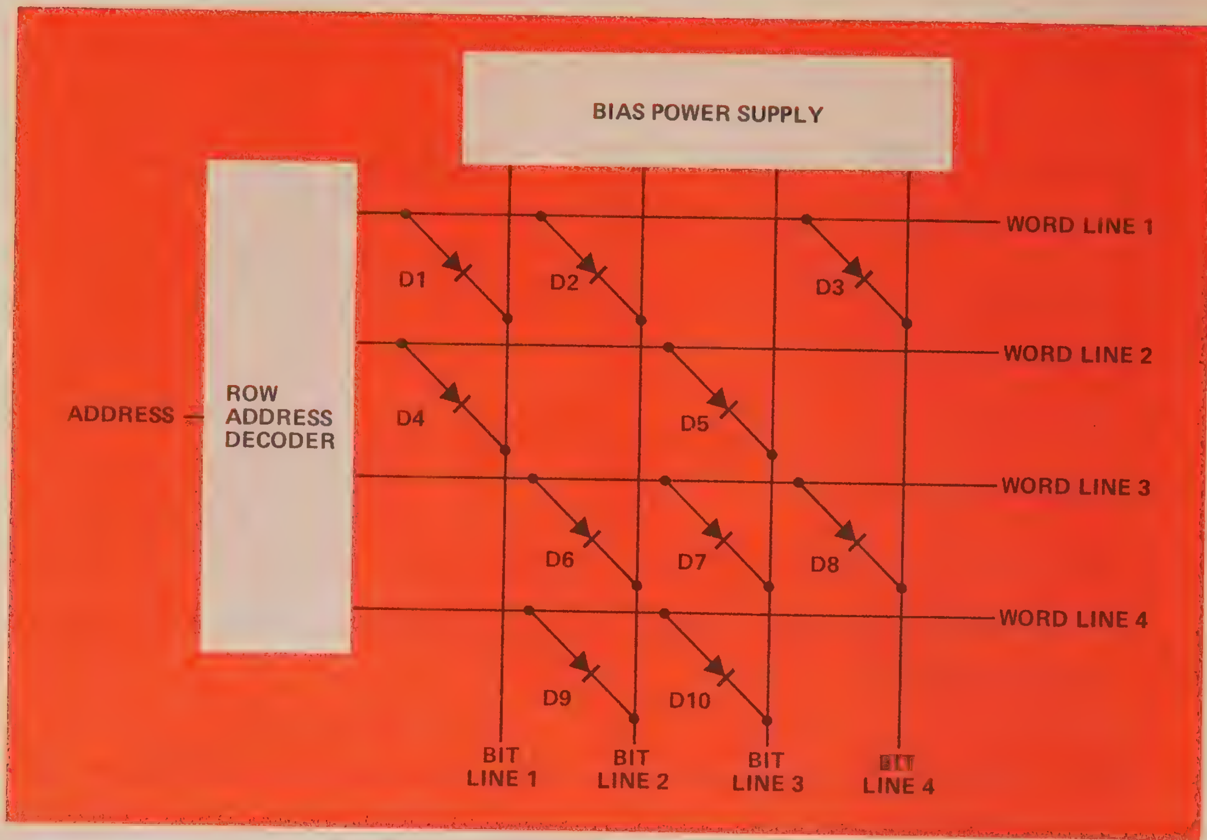


Figure 4. A diode network with random access addressing is the simplest type of semiconductor ROM.

completed, the result is read out and applied to the decoder, which puts it into a form suitable for display.

The basic ROM structure is a matrix of elements, each of which is accessed by a random address code, allowing approximately equal access time to all bits. The simplest ROM structure is a network of diodes wherein the presence or absence of a diode determines the logic state of a particular location; such a network is shown in Figure 4. The row address decoder raises the voltage on the appropriate word line to a high positive level, forward-biasing the diodes attached to that line. When the diodes begin conducting, they force their associated bit lines to a high (logic 1) level, while bit lines not connected to diodes remain low (logic 0). Output amplifiers sense the state of each line and present the bits to the data processor's other circuitry.

For example, if the row address decoder raises the word line 1 level, diodes D1, D2, and D3 conduct, raising bit lines 1, 2 and 4. In this case, the matrix output would be the binary number 1101. The next address might raise word line 4, in which case the output would be 0110. In some applications, column (bit line) addressing is added to select fewer than the maximum possible bit outputs.

ROM matrices are also formed with bipolar and MOS devices. In the most common configurations, the presence or absence of conductors establishes logic states.

Figure 5 shows a ROM matrix utilizing multiple-emitter bipolar transistors. In this case, the collectors are used as row enabling contacts, replacing the word lines, and emitter contacts are omitted from selected locations to set logic levels. When the row address decoder raises the voltage

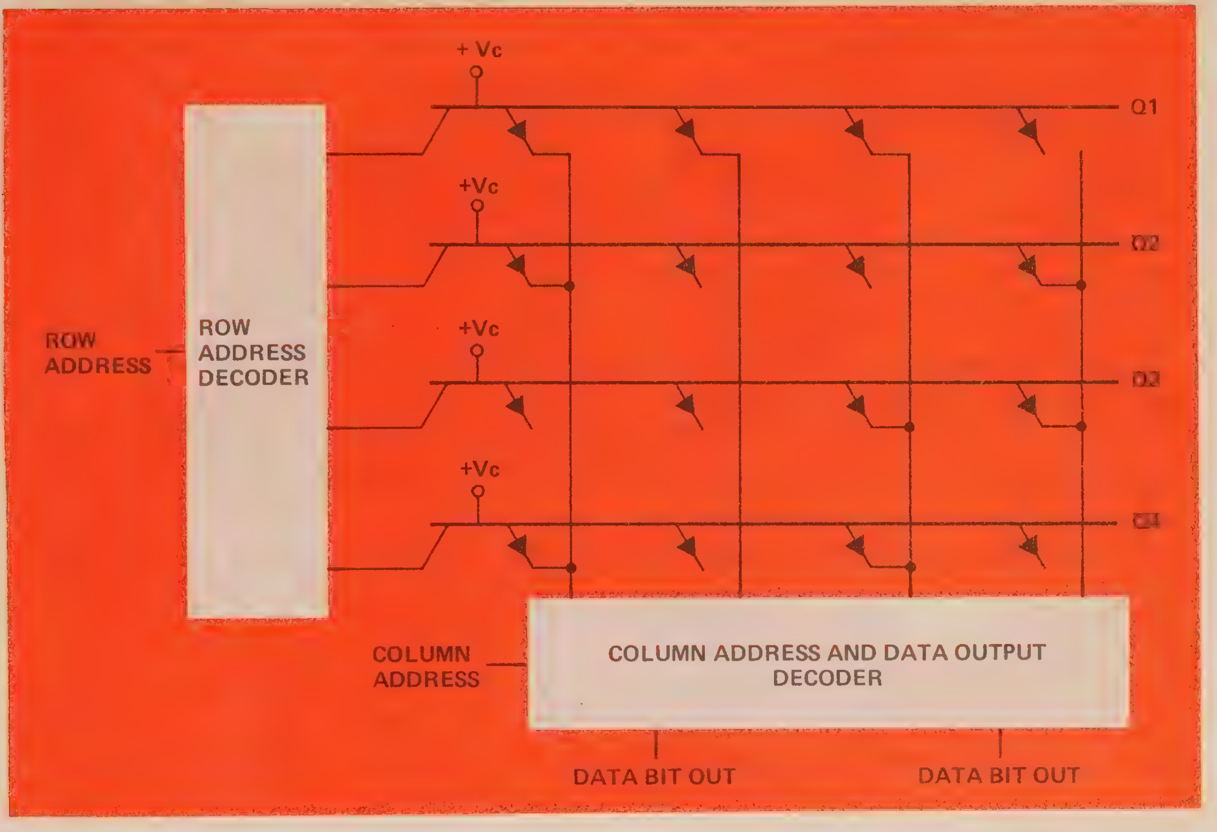


Figure 5. Bipolar read only memory matrix.

on the appropriate collector to a sufficiently high level, the transistor segments with emitter contacts begin conducting; for example, if Q2 is selected, the matrix output is 1001 (the level of columns 1 and 4 raised by conduction, 2 and 3 remaining low). In Figure 5, column address and data output decoding selects two of the four bits for application to the processor.

In Figure 6, a ROM matrix composed of static MOS FET devices is shown. Logic states are determined by the presence or absence of gates within the transistor structures. Reading this memory is accomplished in the same manner as diode and bipolar ROM's, except that the bit lines are driven low (to ground) when the FET's conduct.

Programmable ROM's

Semiconductor memories are almost universally formed on minute silicon chips capable of holding large numbers of integrated circuit devices; the chips often contain complete addressing, decoding, and output circuitry in addition to the memory cells.

In the formation of standard ROM matrices – in which the stored data is never to be changed – logic states are established during the manufacturing process by omitting the proper elements to create the desired bit pattern. This is the most prevalent type of read only memory. There are cases, however, in which standard memories are not available to meet application requirements, so programmable ROM (PROM) matrices are also produced.

A PROM is essentially a semiconductor matrix which has its program written into it at some time other than the manufacturing process. The manufacturer provides a chip on which all of the rows and columns (word and bit lines) are linked by conducting devices. Before integrating the chip into a circuit, the purchaser of the PROM

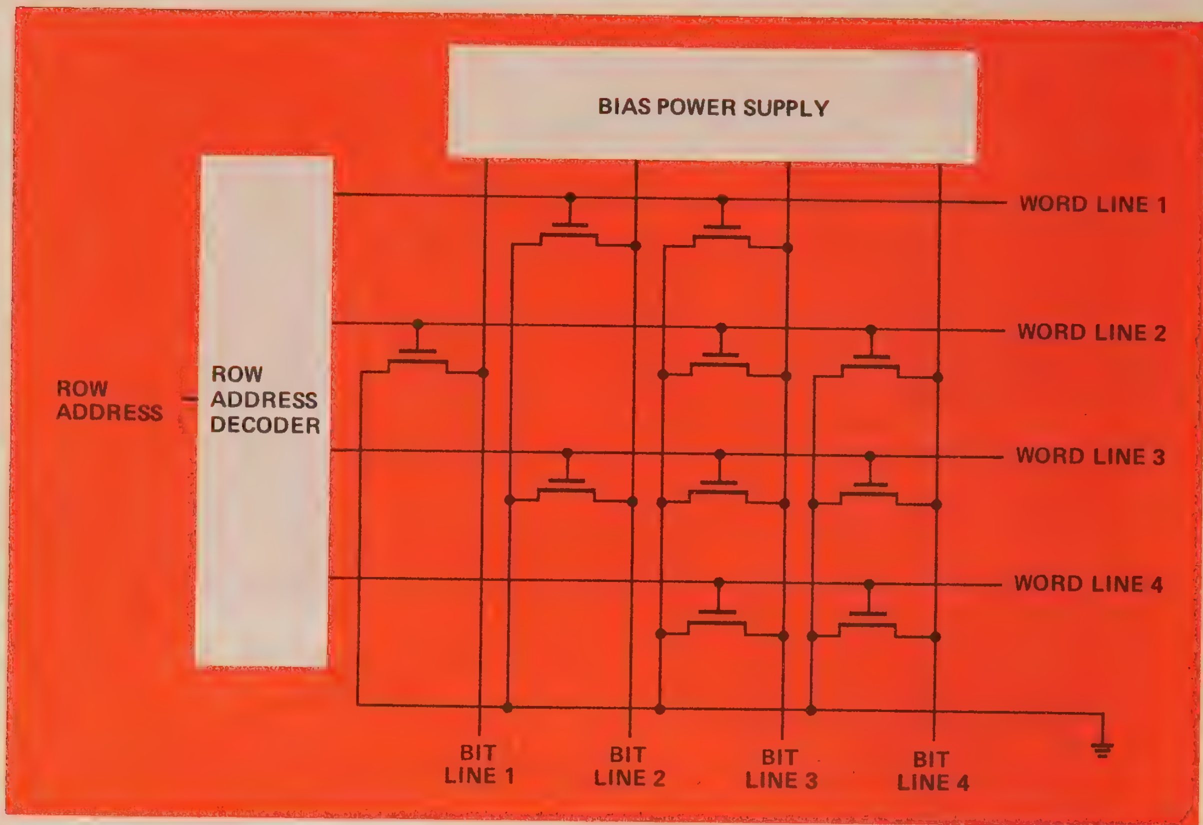


Figure 6. The logic states within a static MOS ROM are determined by the presence or absence of gate contacts.

uses various techniques – from application of a high-level write current to a laser beam – to eliminate devices from the matrix. In this way, a stored program unique to a given application can be produced.

New Developments

This discussion has covered structures that are representative of devices currently used in data processing memory facilities, and has not attempted to consider all of the variations of the basic structures. Advances are being made at a remarkable rate, and today's technology may be totally obsolete in a few years. Among the new memory devices that may bring this about are charge coupled devices (CCD's), bucket brigade devices (BBD's) and magnetic bubbles. Charge coupled and bucket brigade devices are similar in that both store digits as the presence or absence of electric charge.

A basic CCD is a semiconductor chip – either n- or p-type – over which a dielectric material is laid. A series of gate contacts are placed along the dielectric. Charge is stored as minority carriers under the gate regions. When the substrate is p-type, for example, applying a positive voltage to one gate attracts electrons out of the substrate until they dominate in the area directly beneath the gate, forming a "potential well." This storage condition is maintained for times up to several seconds after the gate voltage is reduced. Raising the potential on the next gate in the series forms a second potential well into which the stored charge is transferred; gate potentials are sequentially raised by a clocked voltage to move the stored bit through the device. The CCD is thus a sequentially accessed memory facility similar to the shift register.

The movement of charge in a buck-

et brigade device is the same as in the CCD. Potential wells, however, are replaced by "buckets" of material unlike the substrate; for example, n-type areas may be embedded beneath the gates in a p-type chip to act as MOS storage capacitors.

Fabrication techniques currently limit the production of CCD's and BBD's, but they hold the promise of extremely small, very fast, low-power memories of high density, and many manufacturers are investigating their commercial feasibility.

Magnetic bubble technology is still in the developmental stage, but it also shows great promise. The bubbles, which are tiny, mobile particles whose polarity is opposite to that of the thin film containing them, can be arranged to form coded data patterns, thus providing a storage medium.

Conventional bipolar and MOS devices are being modified to achieve an optimum combination of memory cell size, speed and power consumption. N-channel and p-channel MOS FET's are being combined on one chip as complimentary MOS (CMOS) devices for low-power applications, and Schottky diodes are being introduced into various bipolar configurations to decrease power consumption and increase speed. One such modification has resulted in the low-power Schottky TTL memory cell, which has access speed approaching that of ECL (the fastest presently available cell type) and power requirements close to those of MOS FET's; GTE Lenkurt uses this family of devices in its 262A and 262B data sets to achieve the most rapid data processing possible with the least power. In another development, metal-nitride oxide semiconductors are being looked at as possible non-volatile read/write RAM's (memory facilities which would not lose stored data when power is removed).

Whether improvements to existing structures continue at the present rate, or new technologies take over completely, there is no doubt that semiconductors will play an increasingly important role in data processing systems.

BIBLIOGRAPHY

1. Carr, W. and J. Mize. *MOS/LSI Design and Application*. New York: McGraw-Hill Book Co., 1972.

2. Frankenberg, R. "Designer's Guide to: Semiconductor Memories – Part 2," *EDN*, Vol. 20, No. 15 (August 20, 1975), 58 – 65.

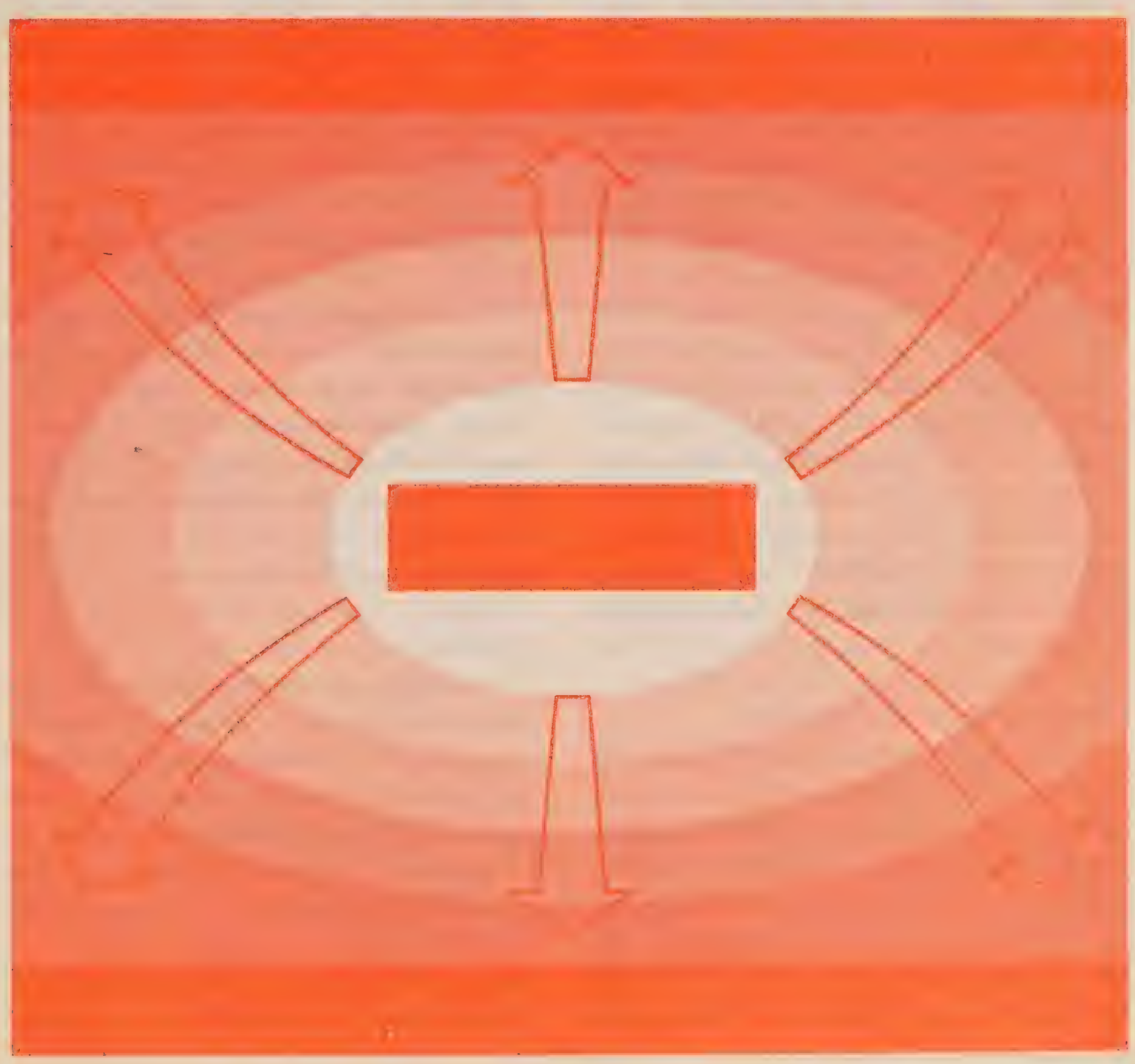
3. Kroeger, J. and B. Threewitt. "Review the Basics of MOS logic," *Electronic Design*, Vol. 22, No. 6 (March 15, 1974), 98 – 105.

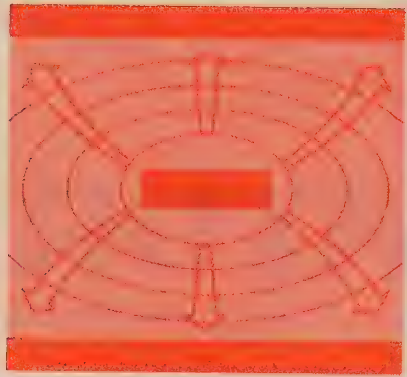
4. Luecke, G. *et. al. Semiconductor Memory Design and Application*. New York: McGraw-Hill Book Co., 1973.

5. Orlicky, J. *The Successful Computer System*. New York: McGraw-Hill Book Co., 1969.

GTE LENKURT
DEMODULATOR
SEPTEMBER 1972

STRIP TRANSMISSION LINES





The transmission line is an important component of any microwave transmission equipment. Within a certain frequency spectrum, strip transmission line provides an efficient and compact conveyance for microwave signals.

A transmission line conducts electromagnetic energy at any frequency from one point or circuit element to another. At low frequencies, transmission of energy is a simple matter, but where microwave frequencies are involved (above 1000 MHz), the problem of transmission becomes much more complex.

The simplest transmission line consists of two parallel wires which are used to transmit low radio frequencies with little electrical loss. At frequencies above two or three hundred megahertz, the losses by radiation of a two-wire line become too high, and coaxial line must be used as the transmission medium. For frequencies above 1000 megahertz, the losses of coaxial line become significant, and waveguide must be used. In recent years, strip transmission line, also known as strip line, and/or flat transmission line, has been developed as a transmission medium for directing electromagnetic energy at microwave frequencies.

Electromagnetic Energy

Electromagnetic energy can be manifested in many forms of radiant energy. Some of these include radio waves, visible light waves, heat waves, gamma rays, X-rays, and cosmic rays. In telecommunications, radio waves are regarded as the main form of radiant energy. When electromagnetic energy in the form of radio waves is propagated along a transmission line, electric and magnetic fields – traveling in a specific direction – are set up around the conductor. This excitation of electric and magnetic fields produces an electromagnetic wave, and it is the energy contained in this wave which the transmission line must convey from one point to another with as little loss as possible. In a two-conductor transmission line, low-frequency electromagnetic energy is propagated along two parallel conductors. Figure 1 shows such a line with its respective electric and magnetic fields.

As an electromagnetic wave propagates along a transmission line, it is attenuated as a function of distance due to resistive losses in the conductors and dielectric losses in the dielectric material. The magnitude of this loss must be low if the transmission line is to function efficiently.

Modes

Closely associated with an electromagnetic wave is its mode of transmission. The electric and magnetic field-configuration in which energy at a given frequency propagates along a transmission line is referred to as a mode. Theoretically, in any transmission line, there is an infinite number of modes which can be excited or established along the line. Each is charac-

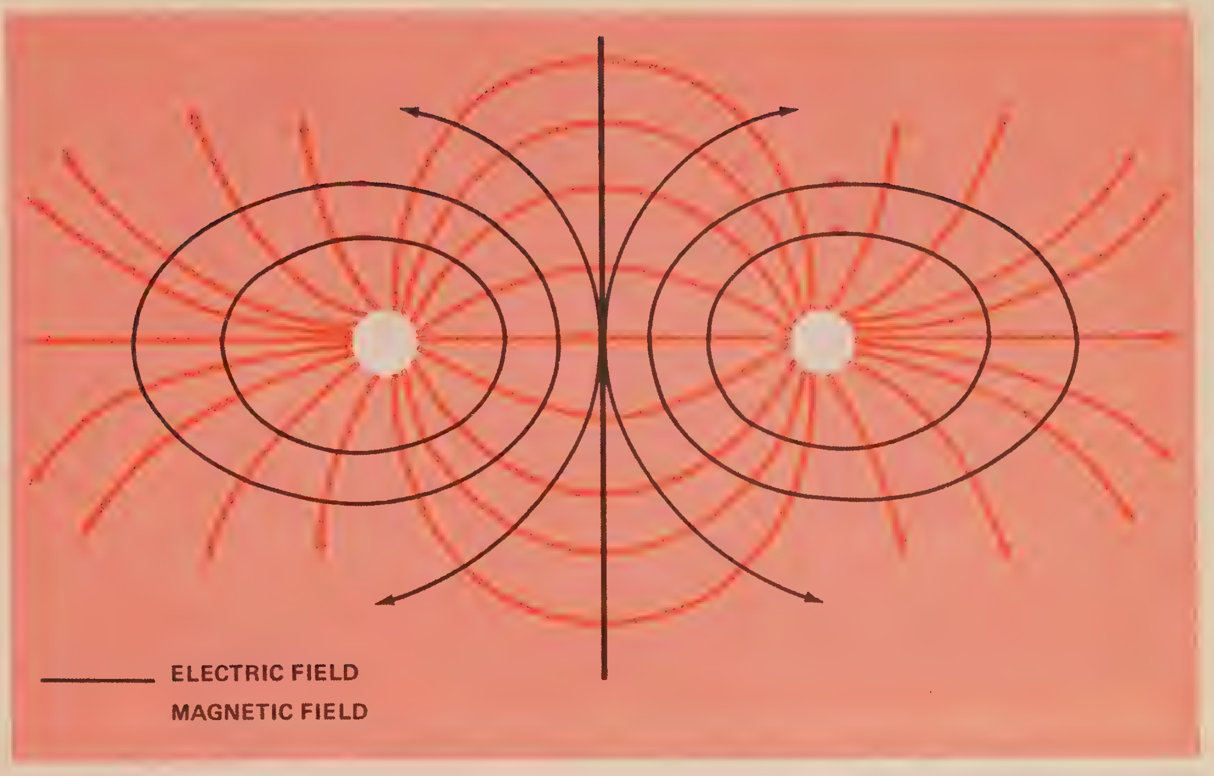


Figure 1. Two-conductor transmission line and its electric and magnetic fields. Here, current enters one line and comes out of the other.

terized by a particular configuration of electric and magnetic fields. In most applications, the line is physically designed so that its operating frequency is capable of efficient energy transmission in only one mode. The appearance of other modes causes an attenuation of the signal, in such a line.

The principal or TEM (transverse electromagnetic) mode is most commonly used with coaxial transmission line. The most outstanding characteristic of this mode is that it consists only of electric and magnetic fields that are transverse to the direction of energy flow. Or, more simply, both electric and magnetic fields are everywhere perpendicular to the direction of propagation. In contrast, higher-order modes have electric or magnetic field components either of which may lie in the direction of energy flow. TEM waves can transmit energy at frequencies down to dc, while the higher-order modes have cutoff frequencies below which energy cannot be transmitted along the line. A transmission line must have two separate conductors to support a TEM wave. (Such as in open two-wire or coaxial transmission line.) Waveguide has only one path of transmission and therefore cannot operate in the principal mode of transmission. It must operate in the transverse electric or transverse magnetic modes. This difference is the essential reason behind the unique modal configurations of coaxial and waveguide transmission mediums.

Coaxial Transmission Line

In coaxial transmission line, electromagnetic waves propagate through the dielectric material which is bounded by the inner and outer conductors. Because of "skin effect" – the tendency of RF current to travel near the surface of a conductor – most of the

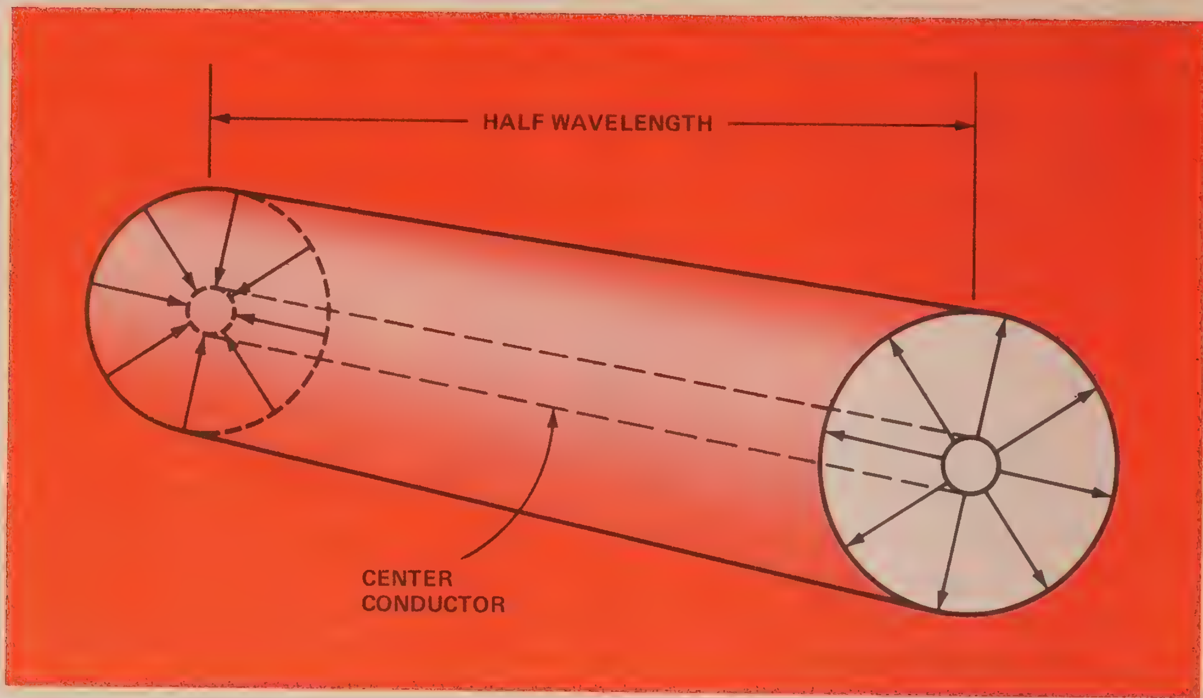


Figure 2. In coaxial transmission line, the electric fields are in opposite directions at each half-wavelength.

electromagnetic current is concentrated on the surfaces of the conductors. The electric fields of coaxial line are all radial in that they are directed toward or away from the conductor at any given time. For each half wavelength along a coaxial line, the electric fields will be in opposite directions as shown in Figure 2.

For the transmission of electromagnetic energy in a single mode, the mean circumference of the coaxial line must be restricted to less than one wavelength. Because of the shorter wavelengths of higher frequencies, the maximum permissible dimensions of coaxial line are small and the losses are increasingly high. The magnitude of those losses may require the use of waveguide or strip transmission lines.

Waveguide

Any transmission line may be referred to as a waveguide since it guides electromagnetic waves from one point to another. More commonly, however, waveguide refers to hollow metal pipes used as microwave transmission lines. They may be rectangular, circular, or elliptical in cross section; but, by far, the most commonly used is rectangular waveguide. In waveguide, electromagnetic waves are propagated through the hollow interior of the metal structure causing electric current to flow along the inner surfaces.

The mode at which electromagnetic energy propagates along a waveguide has a specific cutoff frequency. This is the lowest frequency that will propagate through that transmission line while operating in that mode. The "dominant mode" is the mode which has the lowest frequency of propagation for a set of given waveguide dimensions. It is the mode most often used and appears as shown in Figure 3, at one instant in time. Energy at frequencies below the cutoff frequency is greatly attenuated, while energy above the cutoff frequency is transmitted with very little attenuation.

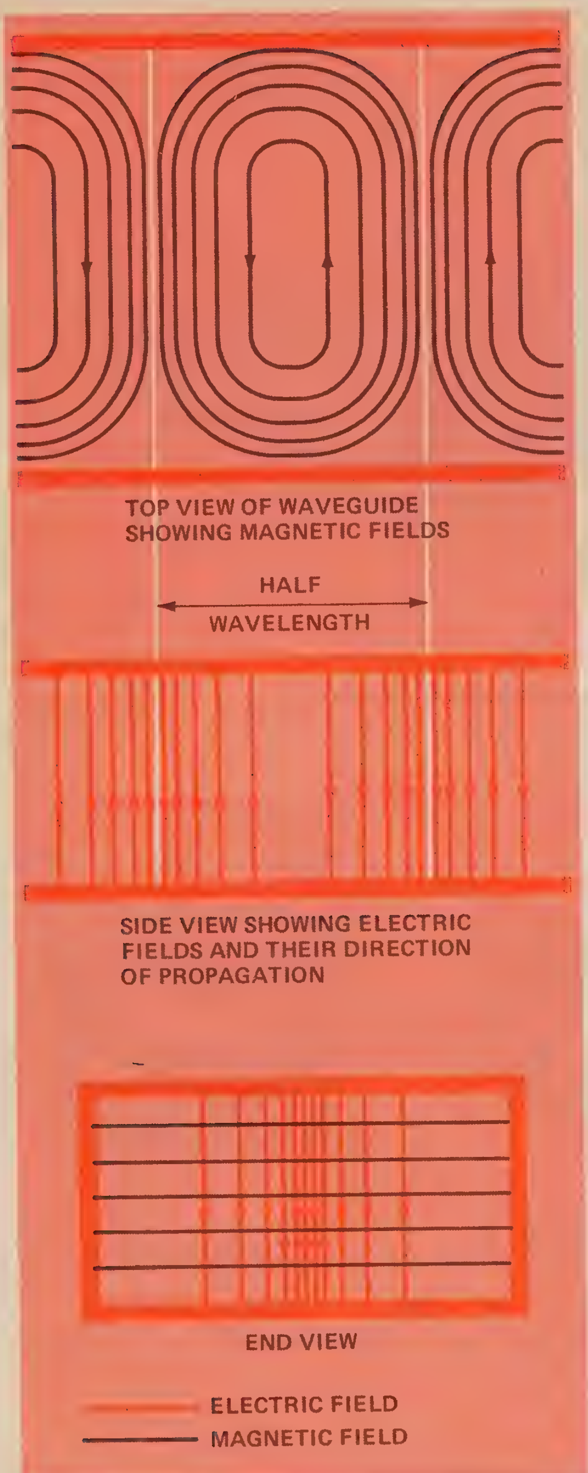


Figure 3. Field configuration of the dominant mode in rectangular waveguide.

For energy to propagate along a waveguide of given dimensions, the operating frequency must be higher than that of the cutoff frequency. For this reason, waveguide becomes too large to be practical at lower frequencies, and its use is therefore limited to the microwave frequency range. In most waveguide applications, the dimensions of the guide are designed so that only one mode of operation will be present at the operating frequency of the system. The greatest advantages of waveguide over coaxial line are higher frequency capacity, lower signal attenuation, and simple structural design.

Strip Transmission Line

A strip transmission line is very similar in electromagnetic properties to coaxial transmission line. It propagates the same kind of wave (TEM) between two conductors as does coaxial line.

The general configuration of a strip transmission line is shown in Figure 4. The strip conductor is bounded by a dielectric material, which, in turn, is confined by two ground planes. The dielectric has the capacity to reduce the physical dimensions of the line for any operating frequency. For example, if a half-wave transmission line is needed in a circuit, strip line will reduce the length of the line by the square root of the dielectric constant. This reduction in length varies with the dielectric material. By selecting the proper material, an efficient and compact circuit may be designed. Compactness, however, is not the main advantage of strip transmission line, its greatest value is in greater ease of manufacture. Once a line has been properly designed to perform a certain function, it can be manufactured in great quantities using standard photoetching techniques. Strip line manufacturing techniques provide precisely-controlled circuitry of lower cost, reduced size and weight, and greater ruggedness and reliability.

The strip transmission line evolved from the coaxial line. Figure 5 shows

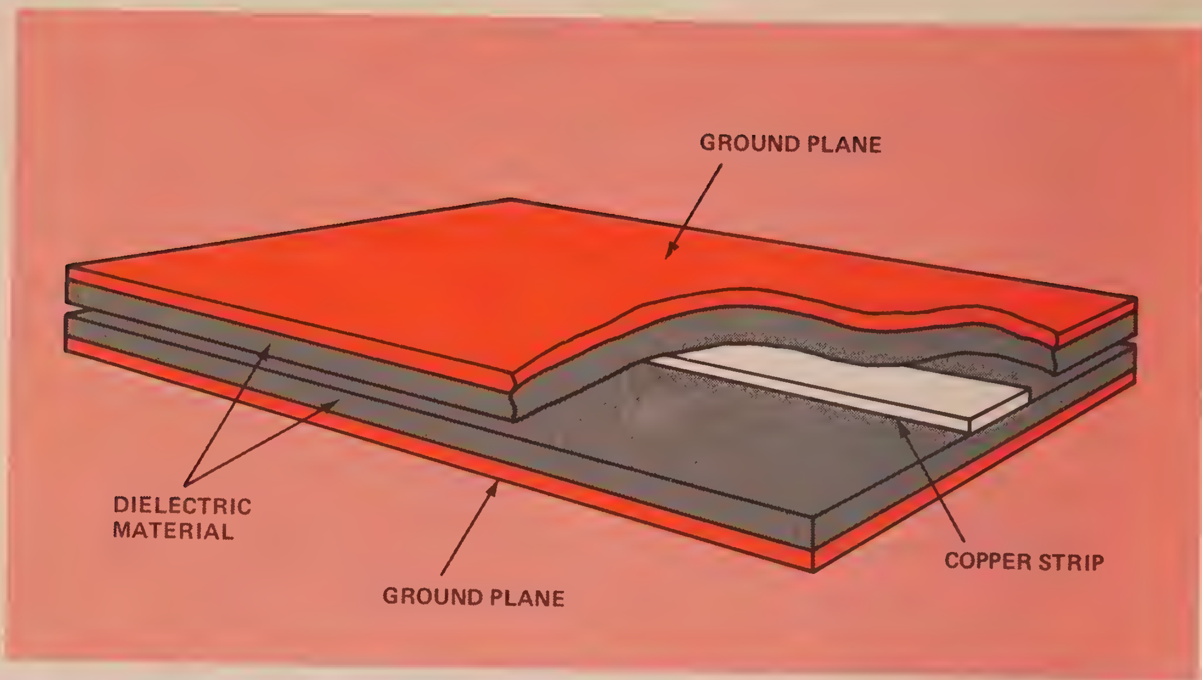


Figure 4. General configuration of a strip transmission line.

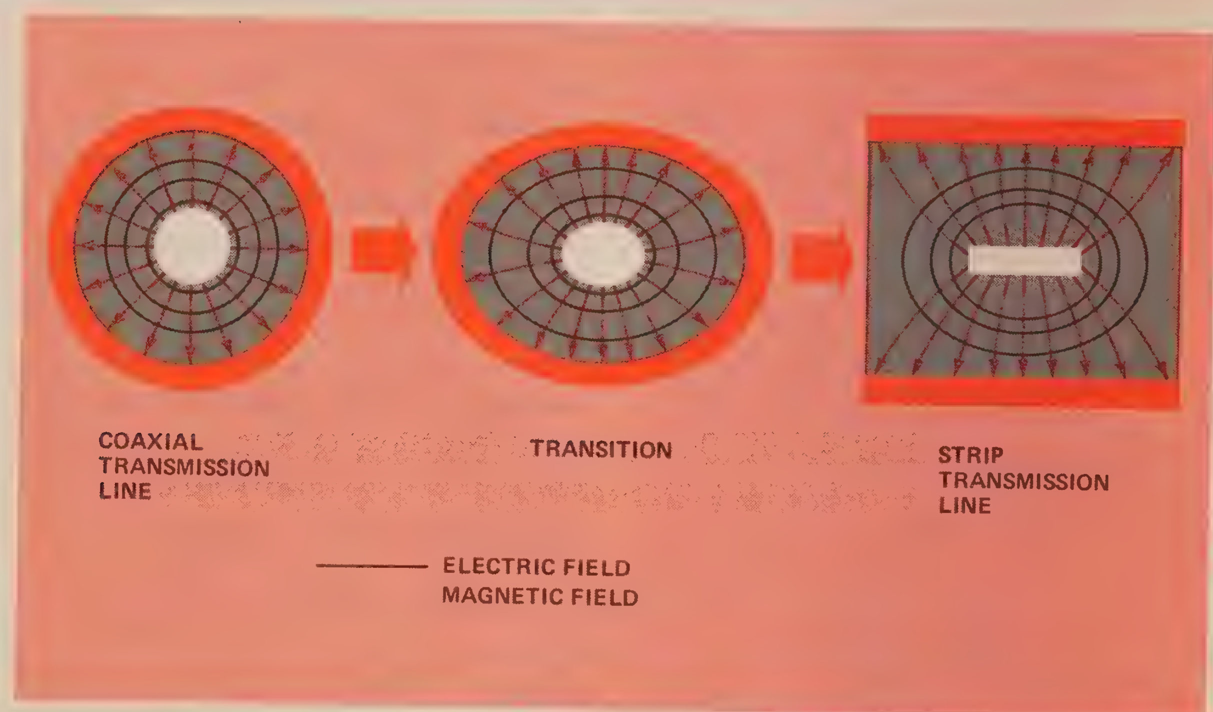


Figure 5. Strip transmission line evolved from the coaxial transmission line.

the evolution from coaxial to strip line with respect to electric and magnetic fields. In coaxial line the magnetic lines of force surround the center conductor while the electric field extends from the inner to the outer conductor. In strip line, the outer conductor divides into two parallel ground planes.

As with coaxial transmission line, the principal mode (TEM) is generally used. The ground planes of the strip line are at the same potential. The field is mainly confined to the area

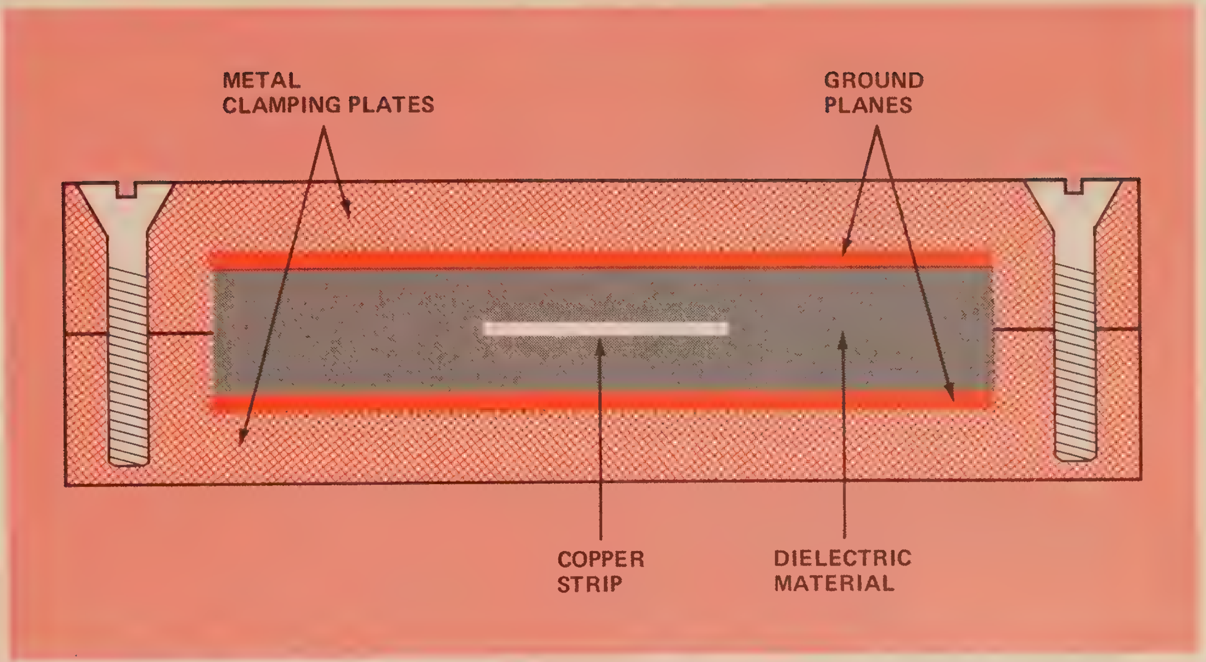


Figure 6. General appearance of a complete strip transmission line. (End view)

between the strip and the ground planes, with the field intensity falling off rapidly with distance from the conducting strip. If the distance from the edge of the strip to the edge of the ground-plane area is as much as two times the spacing between the ground planes, the intensity of the field at the edge is very low.

A wide variety of microwave circuits can be constructed using strip line. Some of these include hybrid Tee's, filters, mixers, circulators, directional couplers, power dividers, antennas, attenuators, and transformers. Such components may be designed and constructed with greater facility than corresponding structures in coaxial transmission line or rectangular waveguide.

Dielectric Material

The dielectric material is an important part of the strip line. It stores the electromagnetic energy which propagates along the conducting strip. The higher the dielectric constant of the material, the greater the concentration of the electric lines of force within the dielectric material.

The energy stored in the dielectric as it travels along is similar to the energy stored in an ocean wave as it moves along the surface of the sea. The energy in the wave is not released until it makes contact with some obstruction such as a ship or a shore. In the case of strip line, the stored energy is not released until it reaches the desired circuit element.

The dielectric constant of a material is a fixed value and is determined by the nature of the material. In recent years it has been possible to effect a desired dielectric constant by mixing materials of different dielectric constant. Such a process is called "dielectric loading" and may, for example, involve the combination of Teflon® (dielectric constant, 2.1) and fiber glass (dielectric constant, 3.88) to give a material with a dielectric constant of 2.37. It is now possible to achieve dielectric constants upwards of 50 by proper mixture of existing materials.

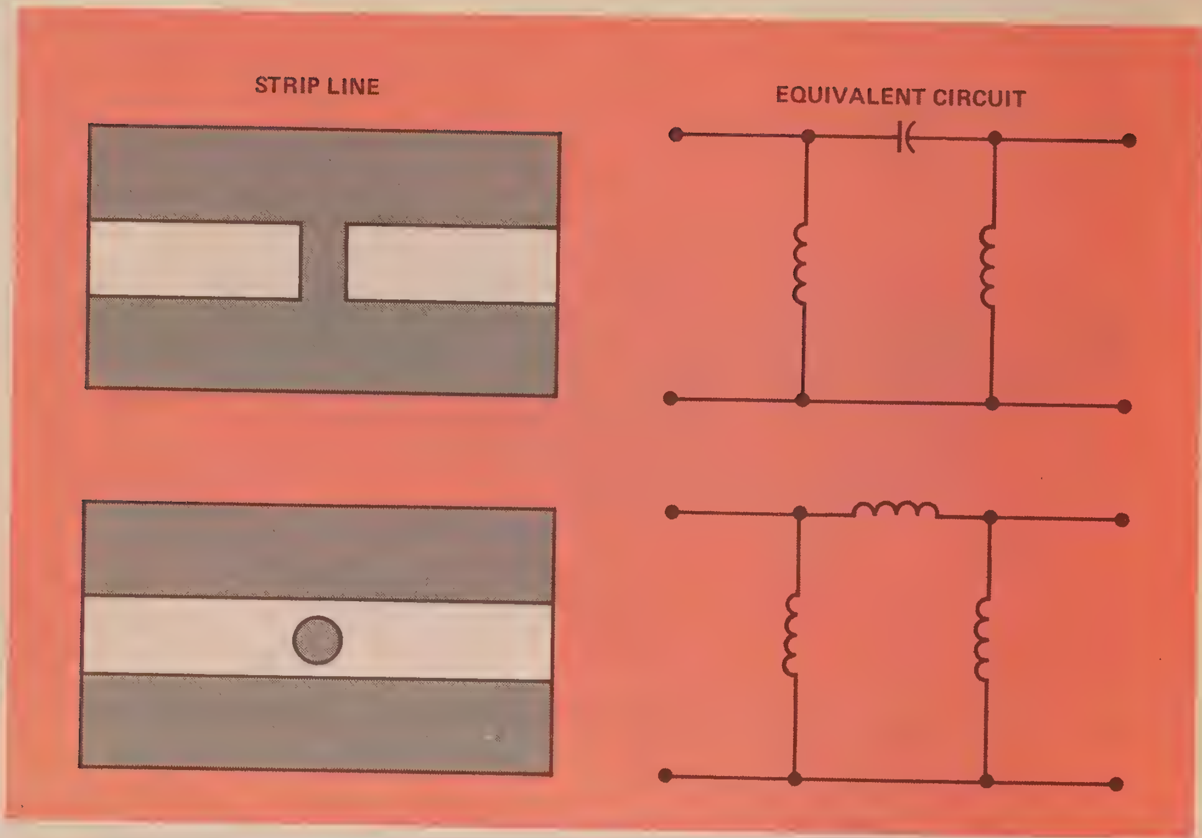


Figure 7. A space or circle positioned along a strip line gives the equivalent circuit of networks made up of discrete components.

Construction

The strip transmission lines designed by GTE Lenkurt are made from two sheets of copper laminates, each of which is copper-clad on both sides, and with the desired dielectric material in the center. On the surface of one sheet, the desired strip is etched to precise dimensions using standard photoetching techniques similar to those used in printed circuit board construction. The second sheet is completely stripped of copper on one side. The two boards are then clamped together to form the complete transmission line as shown in Figure 6. It is also possible to design the strip line with discrete components physically connected to the line and located within a hollow region of the metal clamping structure. This gives an extremely compact device, but it has disadvantages in that additional machining requires extra labor and may also introduce new loss elements.

Unique Properties

Quite often, discussions on strip lines tend to be surrounded by an esoteric aura because of the unique properties and complex physical and mathematical implications involved. In fact, strip transmission lines are still to a large degree under development, but have found many practical uses in the newer microwave equipment.

Although strip lines follow the same electrical laws as does standard electronic circuitry, their physical appearance is markedly different than conventional circuits because of the high frequencies at which they operate. Components such as resistors, capacitors, and coils which are normal-

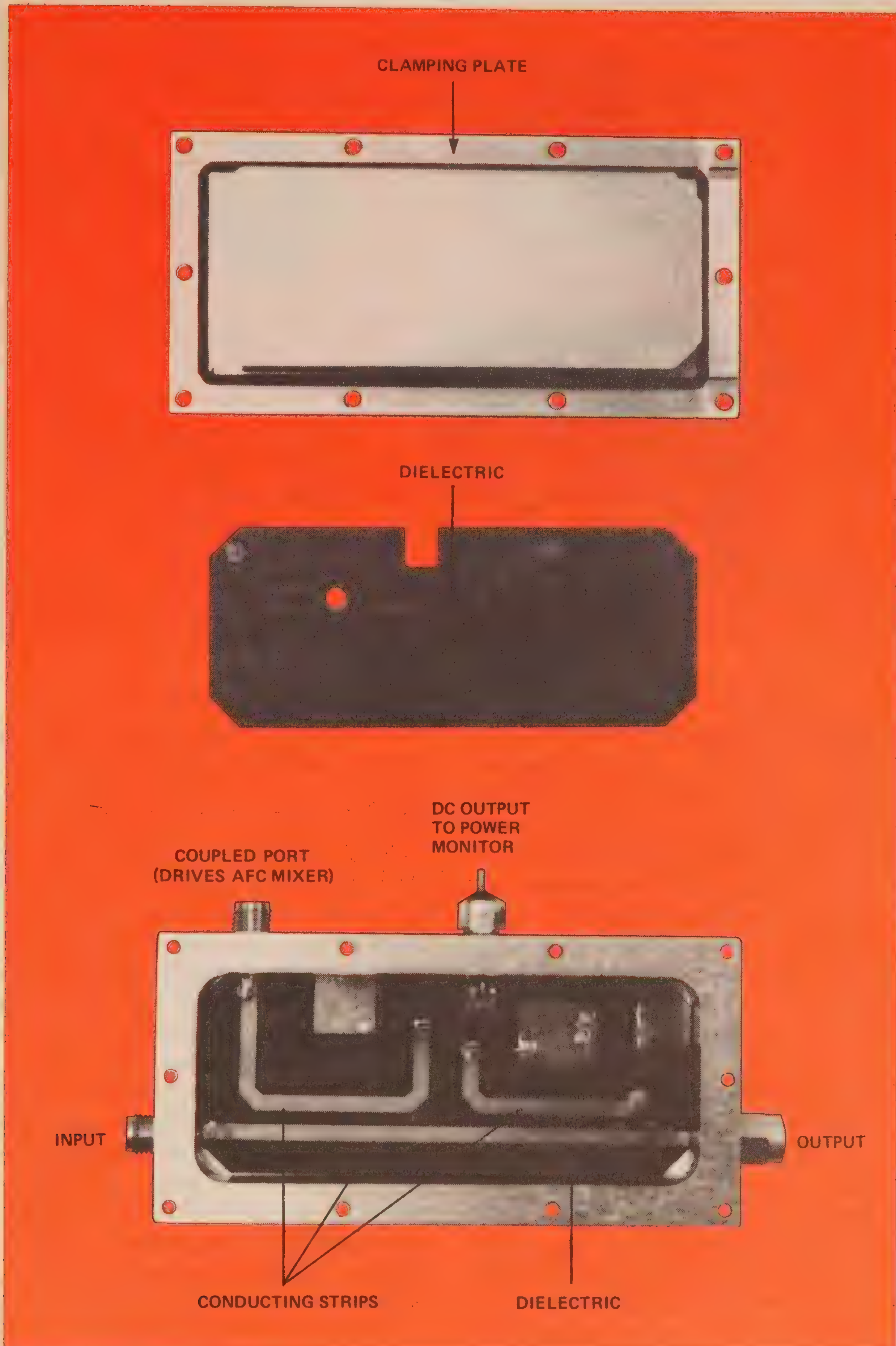


Figure 8. The principal of the strip line is used in construction of directional couplers.

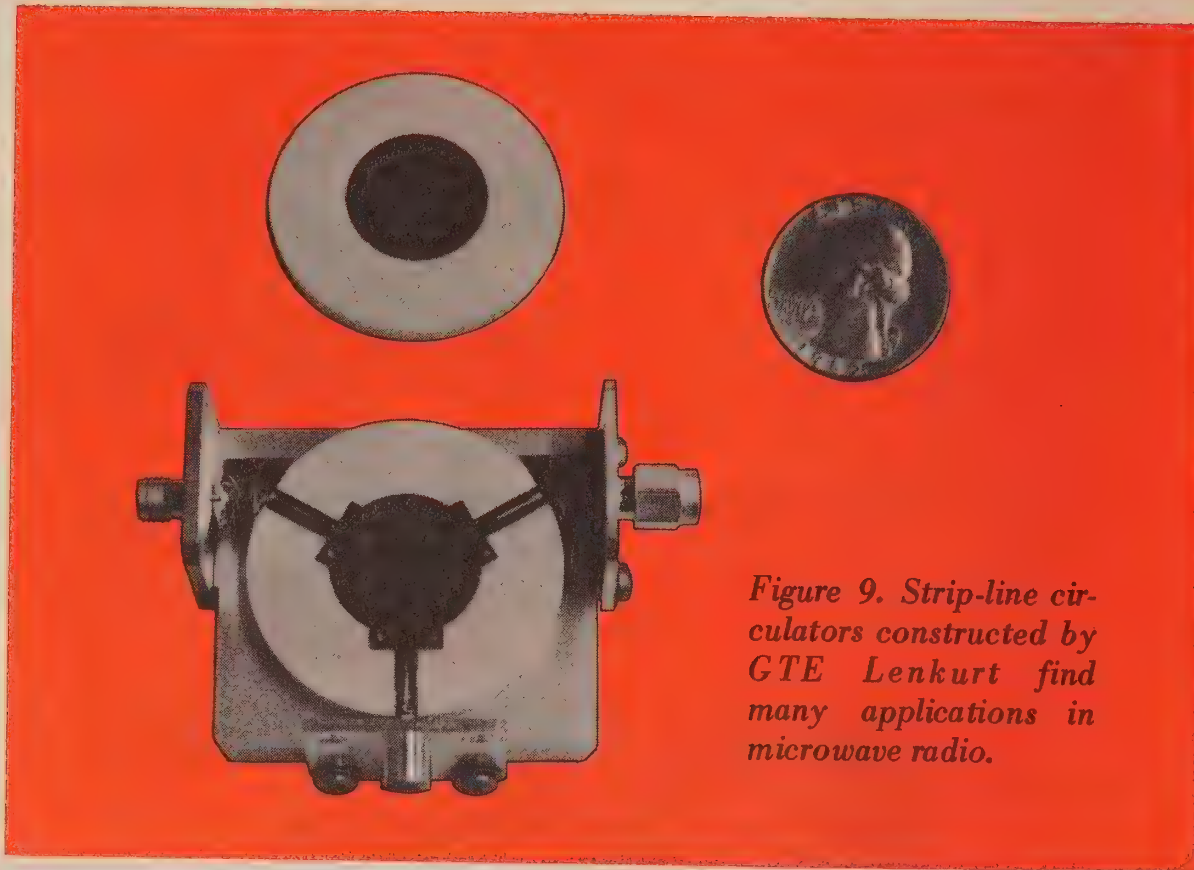


Figure 9. Strip-line circulators constructed by GTE Lenkurt find many applications in microwave radio.

ly used at lower radio frequencies become extremely small and inefficient at microwave frequencies. Microwave circuitry is therefore often manifested in lengths of transmission line rather than discrete components. By designing strip lines of various lengths and patterns, it is possible to achieve results which are equivalent to those obtained by conventional components at lower frequencies.

Microwave Filters

Strip transmission line may be used to provide some of the resonant or reactive elements that make up microwave filters. For example, precise positioning of spaces or holes along the line will give the equivalent result as conventional coil and capacitor networks made of discrete components. (See Figure 7) The inductive and capacitive values of these networks are mainly determined by the length of the space, the diameter of the hole, and the width of the conducting strip.

Strip-Line Directional Couplers

The directional couplers built by GTE Lenkurt in strip-line form utilize more than one conductor on a single dielectric substrate. The physical layout of these couplers adheres to design equations which enable the calculation of the amount of coupling between two conducting strips. The length, width and distance of separation of two strips are crucial factors in determining the amount of energy which can be transferred from one strip to the other. The general layout of a strip line directional coupler with two couplers in tandem appears as shown in Figure 8.

Circulators

A circulator is a microwave coupling device which has a number of

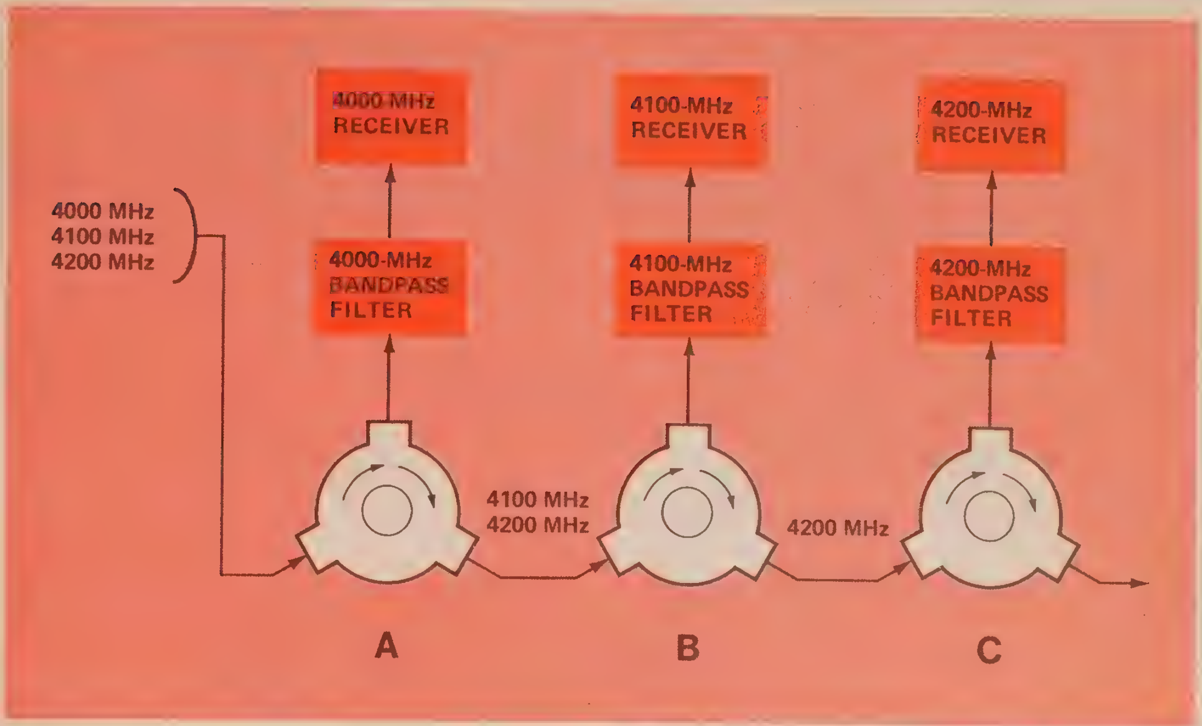


Figure 10. Strip-line circulators may be used as frequency separators for incoming microwave frequencies.

terminals arranged in such a way that energy entering at one terminal is transmitted to the next adjacent terminal in the desired direction. Strip-line circulators such as those built by GTE Lenkurt appear as shown in Figure 9. The circulator uses ferrite magnetic materials in such a configuration that the fields traveling in one direction do not affect the fields traveling in other directions. Circulators find many uses in microwave equipment design. One of their most common uses is in the separation of frequencies of incoming microwave signals as shown in Figure 10. The bandpass filter connected to circulator A may, for example, only accept the 4000-MHz signal. The remaining frequencies will then be transferred via circulator A to circulator B. The bandpass filter connected to circulator B will accept only the 4100-MHz signal, and pass the rest to circulator C. In this way, a whole band of incoming frequencies can be separated and directed to their respective receivers.

The use of strip transmission line has frequency limitations just as coax or waveguide. Some manufacturers utilize strip line at frequencies as high as 8 or 9 GHz, but it is more commonly used at 2 to 4 GHz.

Later Developments

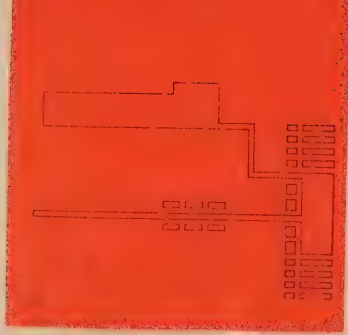
The ease of manufacture and the compactness of the strip line are highly desirable traits, conducive to todays modern microwave components. The strip line principle has brought about other developments in the field of microelectronics. Microstrip circuitry evolved from the strip line, and although some aspects of both techniques are still in the developmental stage, each is finding practical applications and widespread acceptance in the microwave industry.

GTE LENKURT

DEMODULATOR

NOVEMBER 1972

MICROSTRIP TECHNOLOGY



Microstrip transmission lines may be incorporated into microwave equipment design to achieve a compact, integrated unit of lower cost and greater reliability than those composed entirely of discrete components.

The microstrip transmission line is one of several mediums by which electromagnetic waves, traveling at microwave frequencies, can be conveyed from one circuit to another. They may, for example, tie together the stages of the transmit power amplifier in a microwave transmitter. Or, they may be used to connect the circuit elements of mixers and local oscillators in microwave receivers (see Figure 1).

Microstrip transmission line evolved from strip transmission line, which in turn evolved from coaxial transmission line. Like coaxial line and strip line, microstrip line usually operates in the principal or TEM (transverse electromagnetic) mode of transmission. (See the September, 1972 issue of the *Demodulator* for a discussion on principles and applications of strip transmission lines.)

Strip and Microstrip Lines

Strip and microstrip transmission lines are markedly different from each other in construction and material composition. Strip line utilizes a conducting strip between two ground planes and microstrip has a conductor above a single ground plane (see Figure 2). The dielectric constant of a material is an important factor in microstrip operation. It is the degree to which a dielectric material can store electrostatic energy, as compared to air. Only a single ground plane is needed in microstrip because the dielectric constant of the substrate material is much higher than that used for strip line. The concentration of electromagnetic fields between the conducting strip and the dielectric material increases as the dielectric constant of the material is increased, thus allowing less radiation into the air. The high concentration of fields in microstrip circuits permits a more compact overall design than does strip transmission line. But, why use microstrip at all? Isn't it simply another way of doing something that has previously been accomplished by coaxial line, waveguide, or strip line? Part of the answer to these questions is affirmative, but microstrip circuits also have other characteristics which make them particularly desirable in certain applications.

At 1 GHz and higher, microstrip transmission line offers a new approach to microwave circuitry, at a lower cost. Microstrip also provides circuit compactness and greater ease of production than either coaxial or waveguide components. Probably the greatest single advantage of microstrip circuits, from an engineering point of view, is the convenience with which active devices and discrete components can be mounted on the microstrip line. Since the line is open at the top, it is

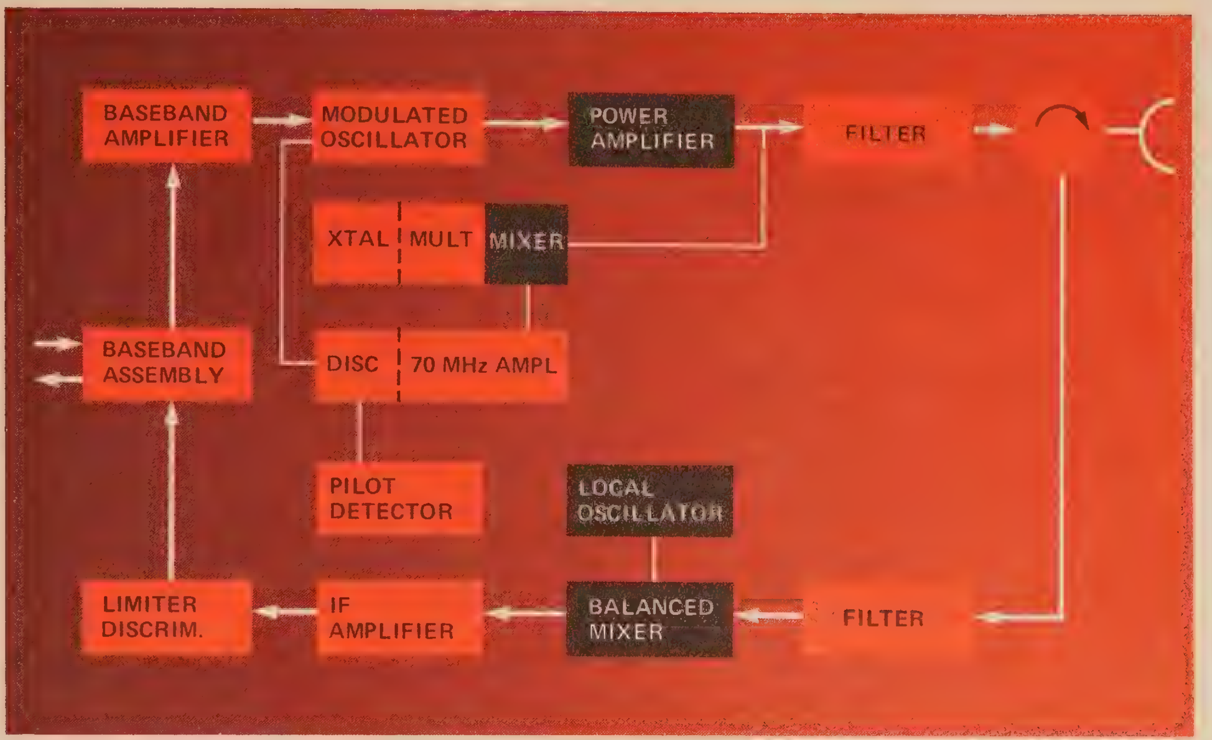


Figure 1. Microstrip circuitry is used to great advantage in some microwave power amplifiers, mixers, and local oscillators.

easily accessible and highly conducive to the mounting of discrete components for both production and experimental purposes.

Microwave Integrated Circuits

When active and passive components are added to a microstrip transmission line, it comes under the general category of a microwave integrated circuit (MIC). The microstrip MIC has a unique geometry for each application because the lengths of microstrip line are of critical importance to its overall function. MIC's are usually of two types – hybrid or monolithic. Hybrid circuits may be constructed by using standard printed circuit photo-etching, and thin or thick-film techniques. The passive components in hybrid circuits, such as resistors, capacitors, and inductors, are manifested in the form of metallic patterns along the conducting strip or soldered on as discrete elements. Active components, such as transistors and diodes, are connected to the passive components in discrete form, thus totally incorporating microstrip transmission line, and active and passive components into one compact package.

Monolithic MIC's contain all active and passive components within the semiconductor substrate itself. To date, the uses for these circuits are somewhat limited in their applications by power restrictions, losses in the substrate, and by difficulties encountered in obtaining a satisfactory yield per production run.

Hybrid Microstrip Circuits

GTE Lenkurt builds hybrid microstrip circuits using the thin-film process, in which passive components are achieved by application of silver, gold, and glass-filled resistive inks. These hybrid circuits often have an alumina dielectric for the substrate, which is a ceramic material with a dielectric constant of approximately 9. The dielectric material decreases the velocity

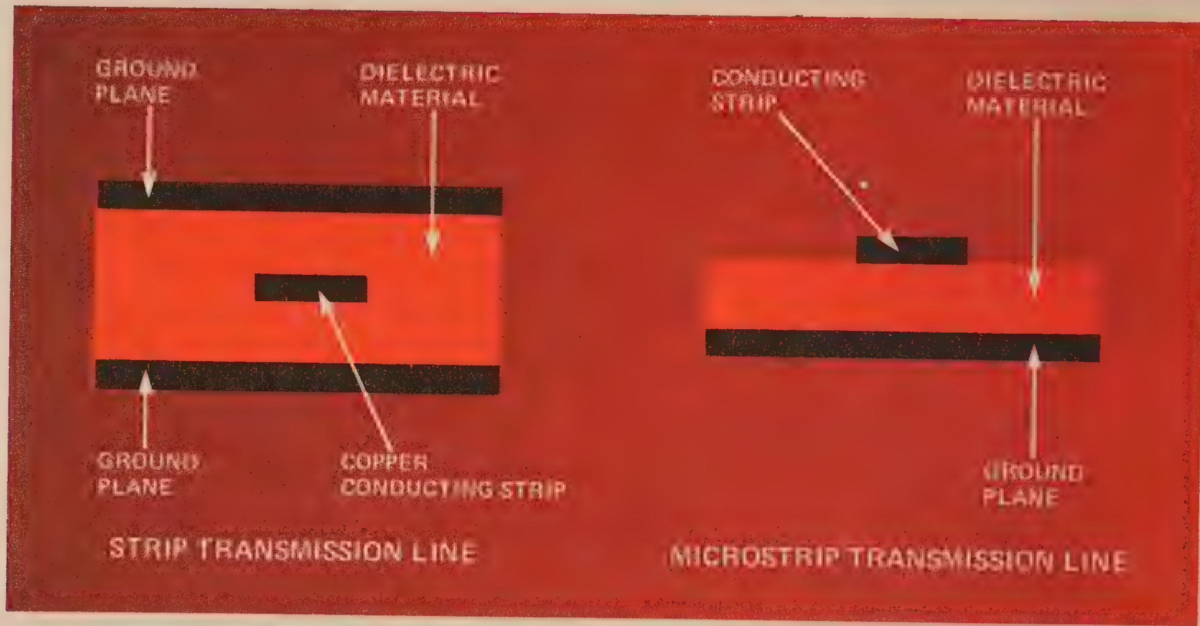


Figure 2. Microstrip transmission lines differ from strip transmission lines in construction as well as in size and type of material.

with which an electromagnetic wave propagates along the microstrip transmission line. This effect reduces or compresses the wavelength to a quantity which is equal to the length of the wave in free space, divided by the square root of the dielectric constant of the substrate material. This is how microstrip circuitry can reduce the physical size of microwave equipment. An alumina substrate, for example, which has a dielectric constant of 9, will reduce the wavelength within the substrate by one-third (wavelength/$\sqrt{9}$).

It is vitally important that the dielectric constant of the substrate material be rigidly maintained during production, since even a slight deviation will cause a change in the electrical properties of the final microstrip circuit. The high dielectric constant and low loss of the alumina substrate permit high-frequency circuits to be metallized on the microstrip transmission line, and at the same time maintain a low loss characteristic. Other materials used for substrates include industrial diamond and sapphire. Diamond is used for high-heat applications, where heat must be transmitted through the substrate and away from the electronic components. Sapphire substrates can be highly polished, which permits very fine lines to be etched on the substrate. This allows the lines to be placed closer together when a high degree of coupling is required in a circuit design. The higher the dielectric constant, the higher the concentration of lines of force within the substrate, and consequently, the more circuits which can be positioned next to each other without danger of interference by radiation. Still, except for military purposes, such as in aircraft equipment, compactness is not the greatest advantage of microstrip circuits. Their greatest economic advantage is in the ease with which they can be mass produced once the correct dimensional and electronic design has been determined.

Manufacture of Microstrip Circuits

GTE Lenkurt's hybrid microstrip circuits begin with an alumina substrate, coated on all six sides with a

Figure 3. The first step in the construction of a microstrip circuit is the drawing of the dimensional pattern at ten times its normal size.

layer of chromium. The layer of chromium is covered by a layer of gold. The chromium layer serves as the adhesive base between the gold and the ceramic, since it has been found that chromium adheres to alumina better than most other materials. The first step in constructing a microstrip circuit is to draw the dimensional pattern of the circuit using a coordinatograph (see Figure 3). This machine is used to draw the circuit ten times its normal size, with a .001 inch tolerance. The drawing will serve to make the film negative which will be imprinted on the substrate. The microstrip circuit board is cleaned ultrasonically, and coated with a photoresistive chemical. The circuit image is then exposed and developed on the substrate. Immersing the substrate in an acid solution dissolves the unwanted gold and chromium, and leaves only the desired image, which has been protected by the photoresist. This process produces circuit boards such as those shown in Figure 4.

The various widths of the metallic strips perform certain electrical functions usually accomplished by passive components in lower frequency circuits. A thick line will produce a capacitive effect, while a thin line will look like an inductance. A gap in the strip functions as a filter, or may act as a coupling transformer. Inductors and capacitors are achieved by the classical approximation of certain lengths of transmission lines which are open or shorted at one end. A useful property

of the microstrip method, then, is that many transmission characteristic impedances can be realized by choosing the correct metallic-strip width. Using the singular characteristics of microstrip transmission lines, the equivalent of lumped components can be achieved.

Soldering and Bonding

Active components are soldered to the circuit using low-temperature indium solder, which, while expensive, has excellent nonleaching properties. Conventional lead-tin solder will leach the gold from the circuit if too much heat is applied, thus changing the design characteristics of the circuit. Laboratory soldering is often done using a hydrogen-flame burner, which quickly heats the metallized surface, since alumina does not absorb infra-red heat. For large-volume production of microstrip circuits, all of the discrete components are soldered onto the substrate simultaneously.

Two relatively new devices used to bond discrete components to the microstrip circuit are the ultrasonic and thermocompression bonders. The ultrasonic bonder vibrates at 60 kHz, exiting the molecules on the leads of discrete components and on the microstrip line, thereby allowing them to merge and adhere to each other. The ultrasonic bonder operates on the principles of adhesion and cohesion. For example, when two highly polished surfaces are brought together, it is very difficult to pull them apart. The two surfaces remain together, not because of an existing vacuum, but because they are molecularly attracted to each other. The molecules are actually sharing their bond strength because of the surface excitation that has taken place during polishing. The 60-kHz vibration of the ultrasonic bonder uses this same principle to weld connections on microstrip circuits.

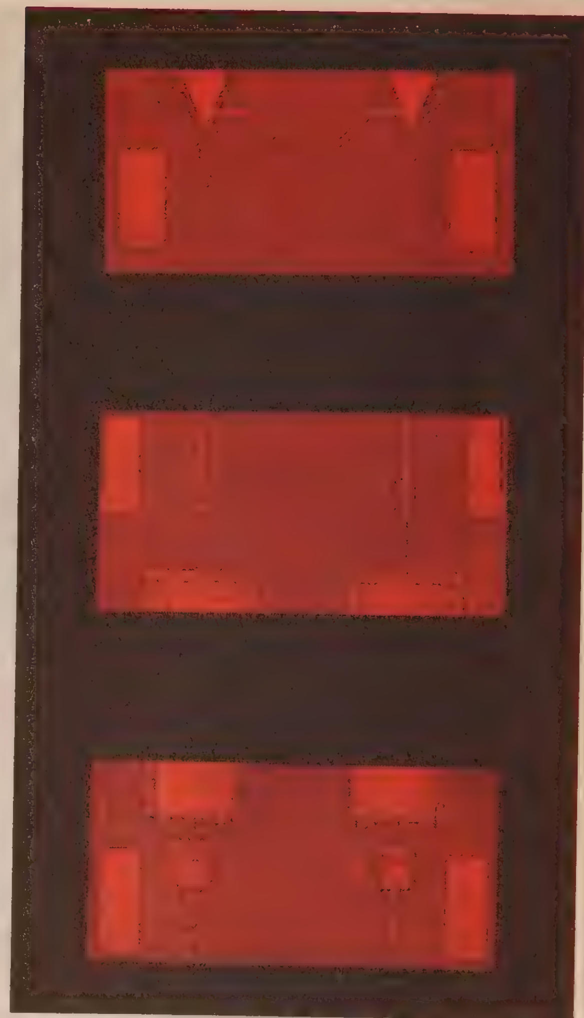


Figure 4. After etching, microstrip substrates are ready for the mounting of discrete components.

The thermocompression bonder heats the component leads and the gold circuitry, under pressure, until they are fused together. This process takes about half a second.

One disadvantage encountered in designing microstrip circuitry, is a certain loss in design time. The engineer cannot accurately construct the circuit in breadboard fashion to see if it will function properly. He cannot change the values of some components as easily as could be done in lower-frequency circuits. Each trial design must be painstakingly engineered and fabricated as if it were the final mold from which all other devices were to be cast. The cost in time, however, of design-

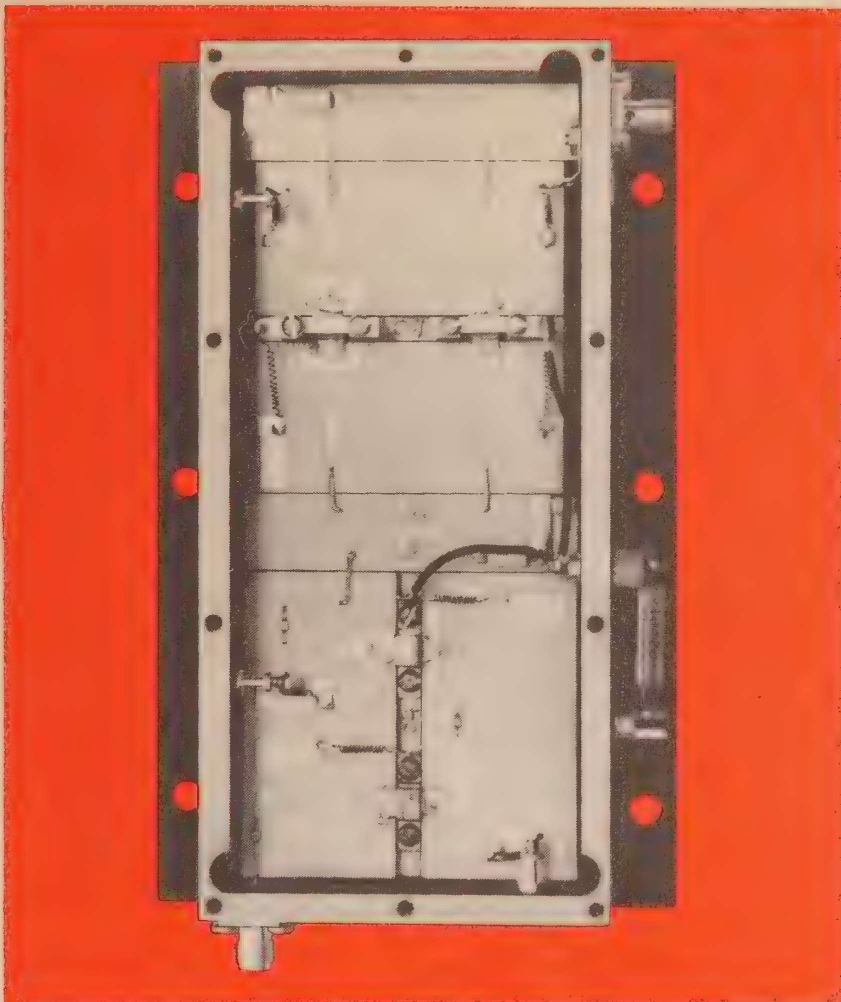


Figure 5. Hybrid microstrip circuits in a microwave power amplifier.

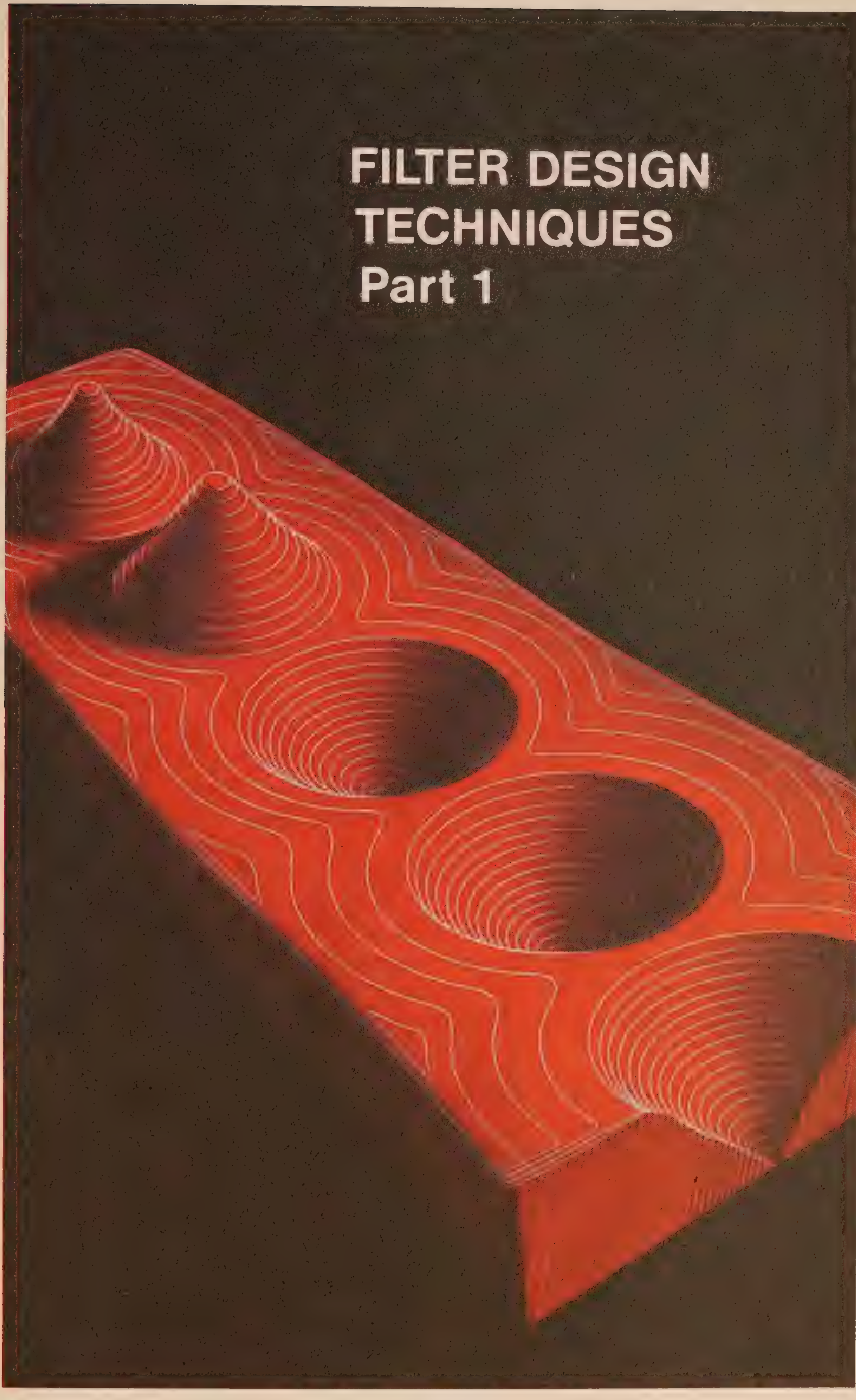
ing and building a microstrip circuit may be offset by the greater ease in manufacture and the higher reproducibility of the final product.

Microstrip circuitry is used only where it has a clear advantage over other transmission mediums. Figure 5 shows extensive use of microstrip circuitry in the 2-GHz power amplifier of a GTE Lenkurt 78C2 microwave radio. The tuning of such a circuit is performed by application or deletion of silver paint at certain areas along the substrate. This will alter the component value of the circuit, and give the amplifier the desired output characteristics.

Much experimental work still remains to be done with strip and microstrip transmission-line circuits, both in the development of monolithic and hybrid techniques, and in the extension of the use of microstrip circuitry to new electronic applications.

Microstrip circuitry should not be regarded as a panacea which will immediately solve all microwave transmission line problems. At this time, many transmission applications are more easily handled by using strip line, coaxial line or waveguide, and it is unlikely that microstrip will ever totally replace any of these mediums.

The telecommunications industry is well along in finding uses for microstrip circuitry, and it appears that future equipment will utilize an increasing amount of this circuitry, but only where it has definite advantages over other existing forms of transmission mediums.


FILTER DESIGN
TECHNIQUES
Part 1

Filters in modern technological applications serve to direct, channel, divide, integrate, and transform, but in all of these functions the basic principle remains the same: selective processing of electrical energy and signal information.

Filters are essentially devices with particular transmission characteristics which can be designed to accept or reject – to selectively process – portions of the electromagnetic frequency spectrum. In this, filters behave like interconnected tuned circuits, which are combinations of resistance, inductance, and capacitance whose reactances produce unique frequency responses.

When an alternating current source is connected to a tuned circuit, the combined effect of capacitive reactance (X_C), inductive reactance (X_L), and resistance (R) constitutes an impedance (Z) – an opposition to the flow of alternating current – which varies with the frequency of the applied source. At low frequencies, X_C is much greater than X_L or R, while at high frequencies, X_L is the largest element. At some intermediate frequency (f_o), X_C equals X_L, and the circuit is said to be resonant, or in a state of resonance.

The effect of the resonant state on a tuned circuit depends in great part upon the manner in which the components are connected. The impedance of a series tuned circuit appears as a large inductive or capacitive reactance at frequencies above and below resonance, thus limiting the current flow in the circuit. At resonance, the effects of X_L and X_C cancel one another, leaving the impedance at a minimum level – equal to the value of R – and allowing maximum current flow (see Figure 1a).

At frequencies below resonance in a parallel tuned circuit, maximum current passes through the low reactance of the inductive component, as at high frequencies current passes readily through the low reactance of the capacitive element. At resonance, however, the reactances are equal, resulting in the lowest level of current through the circuit (see Figure 1b).

While a tuned circuit produces an amplitude peak or trough at a single resonant frequency, passive filter networks – which are also treated as combinations of resistances, inductances, and capacitances – provide constant transmission for a range of frequencies (the passband) and a high degree of attenuation to all other frequencies (the stopband). A low-pass filter, for example, is intended to reject all signals above a specific cutoff frequency; to accomplish this, it produces increased attenuation to the signal as the source frequency is increased (see Figure 2), much as impedance increases in a parallel-resonant tuned circuit as resonance is approached. A high-pass filter, on the other hand, is designed to reject all signals

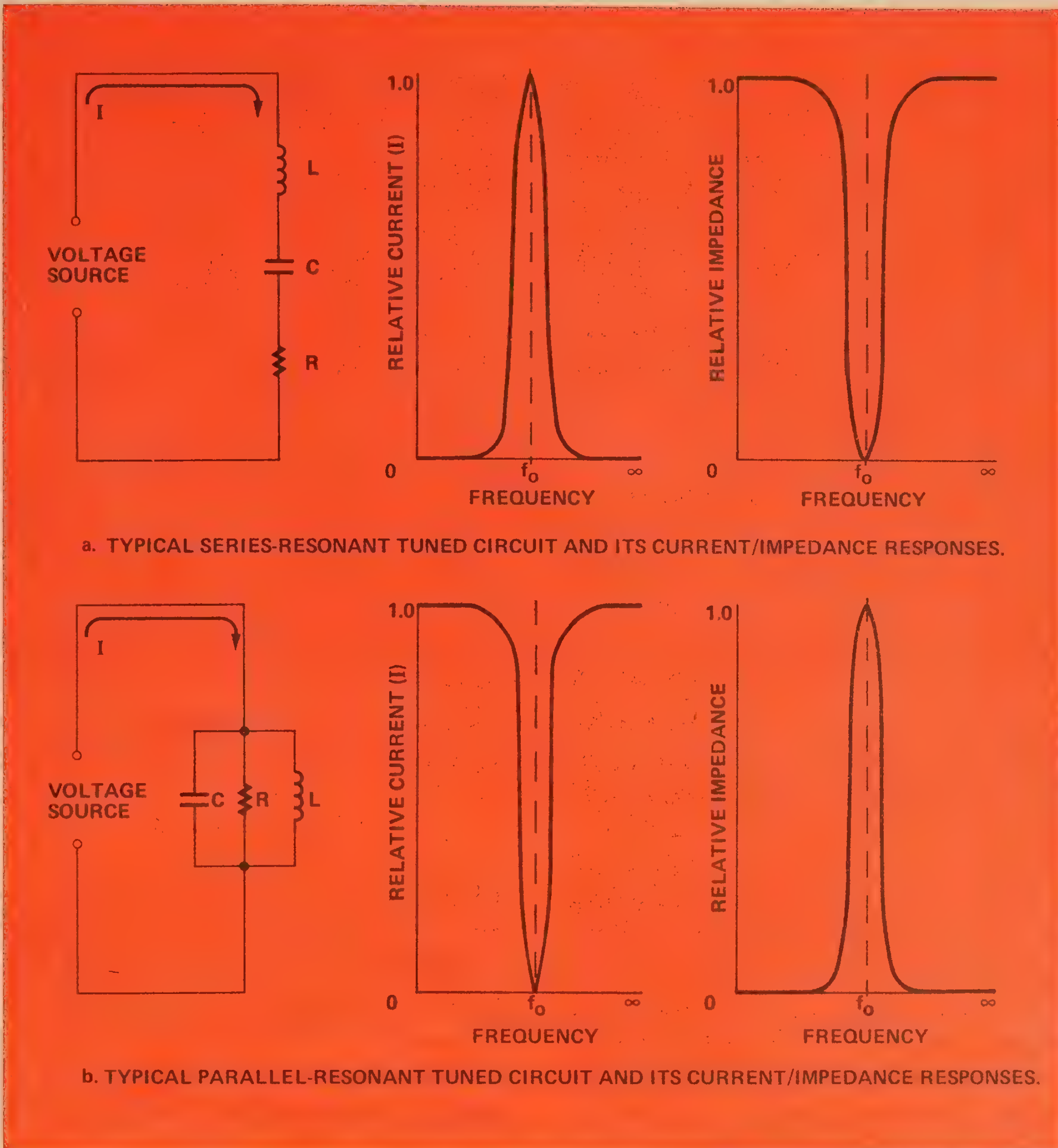


Figure 1. The interaction of component characteristics causes a peak or trough in the current/impedance response of a resonant circuit at one frequency (f_o). Theoretically, the maximum amplitude achieved by either element is infinity.

below cut-off, so its attenuation decreases with frequency, as the impedance of a series-resonant tuned circuit decreases when resonance is reached.

Thus, one way to view filters is as a collection of resonant circuits, the sum of whose resonance characteristics determines the overall performance of the network. This is essentially the approach taken in the image parameter theory of filter design.

Image Parameter Theory

The image parameter approach to filter design grew out of the study of transmission lines in the early days of electromagnetic communications. At that time, long lengths of wire were used to send telegraph and telephone signals, so it was natural that they should become the focus of considerable study. It was soon recognized that the action of a transmission line, re-

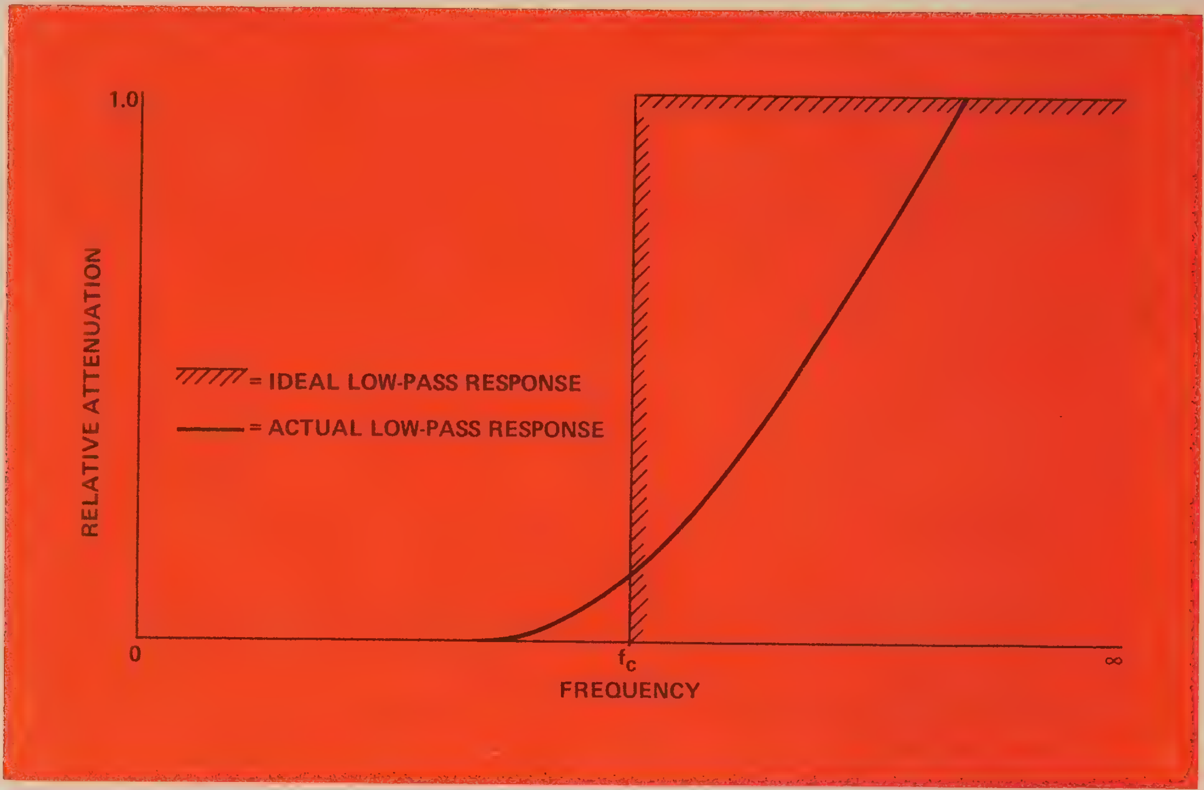


Figure 2. Low-pass filters reject signals above a specific cut-off frequency (f_c) by providing increased attenuation. The ideal is instant rejection at f_c; the actuality is a curve increasing with frequency.

gardless of its length, can be explained in terms of resistance, reactance, and impedance, although lumped physical components may not be present. Dividing such a line into successively smaller segments, a basic unit is eventually arrived at which can be described as a network with two input terminals and two output terminals, with each pair of terminals defining a port (see Figure 3). This two-port transmission network, which is one of the most common types of simple filter, possesses transmission and impedance properties which depend upon two quantities: image impedance and image transfer function.

Image Impedance

Image impedance is the filter network equivalent of the characteristic impedance of a transmission line.

The characteristic impedance, Z_o, of a line is that opposition to ac which, when used to terminate one end of the line, causes the input impedance to be of the same value. Image impedance, Z_i, is actually an extension of this concept, taking into account the non-symmetrical nature of many filter networks, where the im-

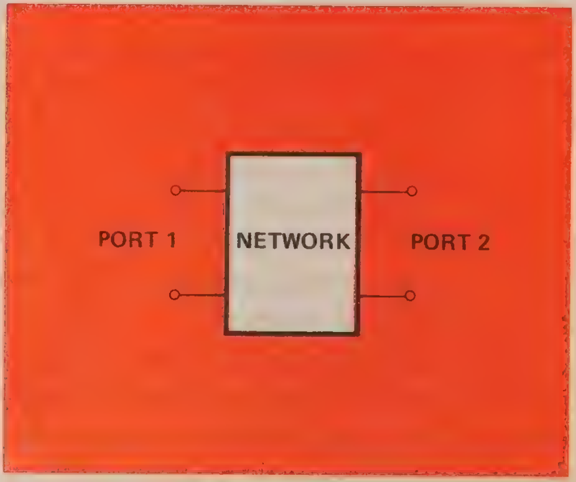


Figure 3. A four-terminal network has two ports, either of which may be used as an input point.

pedance looking into one port may differ from that looking into the other. In such asymmetrical networks, there are two image impedances, one for each port. Where a Z_o termination of a transmission line causes a Z_o input impedance, terminating one port of a filter network causes the other port to exhibit its own Z_i as the input impedance; that is, one Z_i appears as the image of the other (see Figure 4). A symmetrical network closely resembles a transmission line, in that the image impedances of both ports are equal (terminating one port in its Z_i results in the same Z_i appearing as the input impedance).

The image impedance of a given port can be defined as the geometric mean of the network short and open circuit conditions. That is, in a two-port network, the image impedance of port 1 is expressed as

$$\sqrt{\frac{Z_{11}}{Y_{11}}}$$

and the image impedance of port 2 is

$$\sqrt{\frac{Z_{22}}{Y_{22}}}$$

where Z_{11} is the impedance presented to an input signal at port 1 with port 2 open, and Y_{11} is the admittance (inverse property of impedance) presented at port 1 with port 2 shorted. Z_{22} and Y_{22} express the same properties when the input is at port 2 and port 1 is open or shorted. Image impedance, then, approximates the impedance of a network in a state half way between short and open, and represents an average figure for the range of conditions between the two extremes.

As the basic filter unit in image parameter design, the half-section is the smallest two-port combination of elements exhibiting an image impedance characteristic. In image parameter filter design, half-sections are combined in cascade to produce the required filter response. In connecting the output terminals of one half-section to the input terminals of the next (cascade connection), it is necessary that the image impedances be matched through the length of the filter. In this way, maximum power transfer can be attained at passband frequencies, where the image impedances appear.

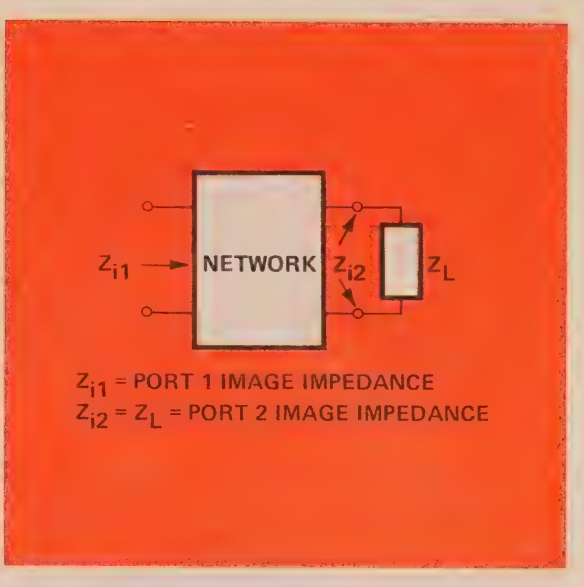


Figure 4. When one port of a two-port network is terminated in its image impedance, the image impedance of the other port becomes the network input impedance.

For example, the half-section for a basic low-pass constant-k filter – in which the product of the impedances is independent of frequency – would consist of a series inductance and a shunt capacitance (see Figure 5a), producing a stopband with attenuation increasing to infinity with frequency, as shown in Figure 2. To make a full low-pass filter section, two half-sections are joined so that their image impedances are properly matched (see Figure 5b); successive sections can be added to provide whatever response characteristic is required.

The m-derived half-section is similar to a constant-k unit with a resonant

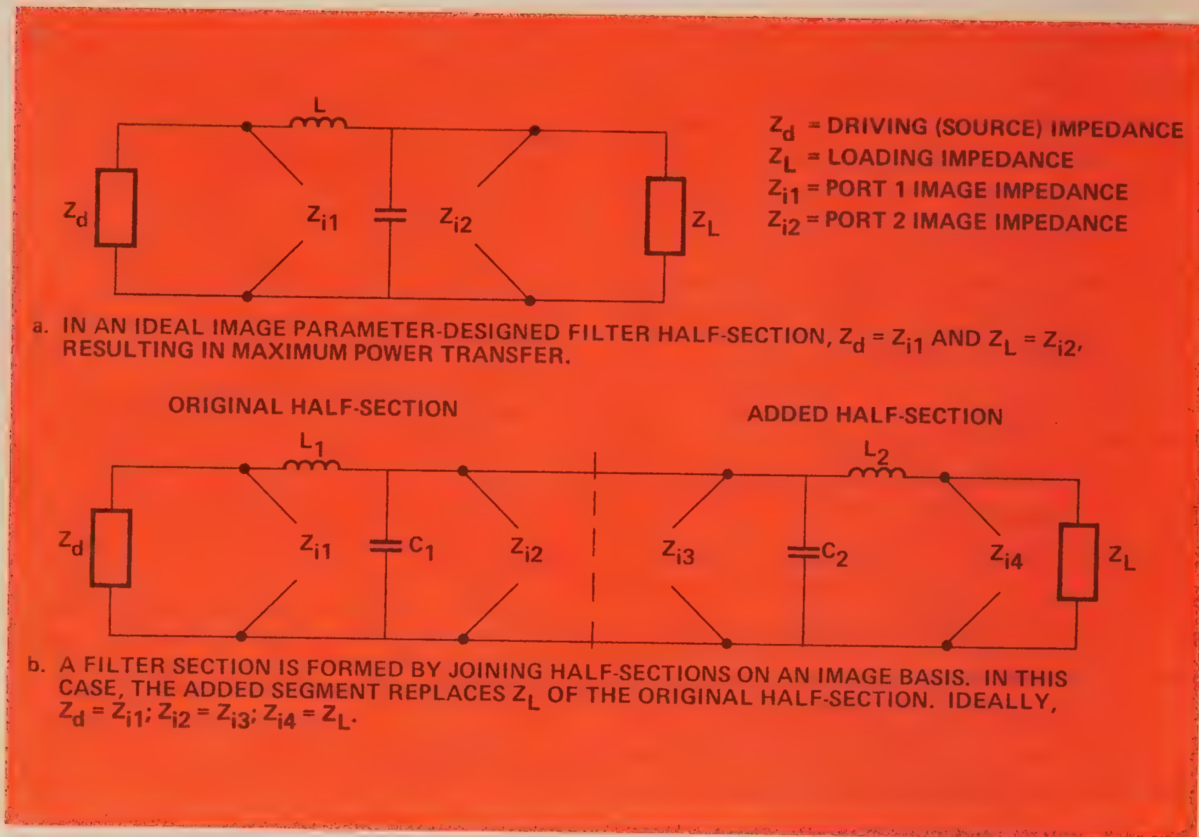


Figure 5. In image parameter filter design, it is assumed that the various half-sections are terminated in their image impedances.

segment introduced (see Figure 6a). The stopband of this filter has a peak at the resonant frequency of the tank circuit, with the rest of the band exhibiting constant-k type characteristics (see Figure 6b). The advantage of the m-derived filter is that several m segments can be added and the attenuation peaks – which theoretically go to infinity – can be placed anywhere in the stopband, providing extremely sharp cutoff points and increased attenuation at frequencies where the circuit design requires them. This allows the filter to produce only the stopband attenuation levels required, eliminating the components that would be necessary if infinite attenuation were sought at all stopband frequencies.

Image Transfer Function

Closely related to the image impedance of a filter is the image transfer function (τ), which is a measure of the response of a two-port network when it is terminated at both ports – driven and loaded – by its image impedances. The image transfer function consists of two terms: an "image attenuation constant" (α), and an "image phase constant" (β). These elements are related by the expression $\tau = \alpha + j\beta$ (where $j = \sqrt{-1}$), so that if both ports of the network are terminated by their image impedances, the attenuation and phase shift seen in the output constitute the image transfer function.

When a series of half-sections are joined to form a more complex filter, and the filter is terminated at both ports by impedance corresponding to the Z_i of the network, then the resultant impedance appears nominally resistive to passband frequencies. In this case, power is absorbed from the source and transferred to the load, with the amount of transference de-

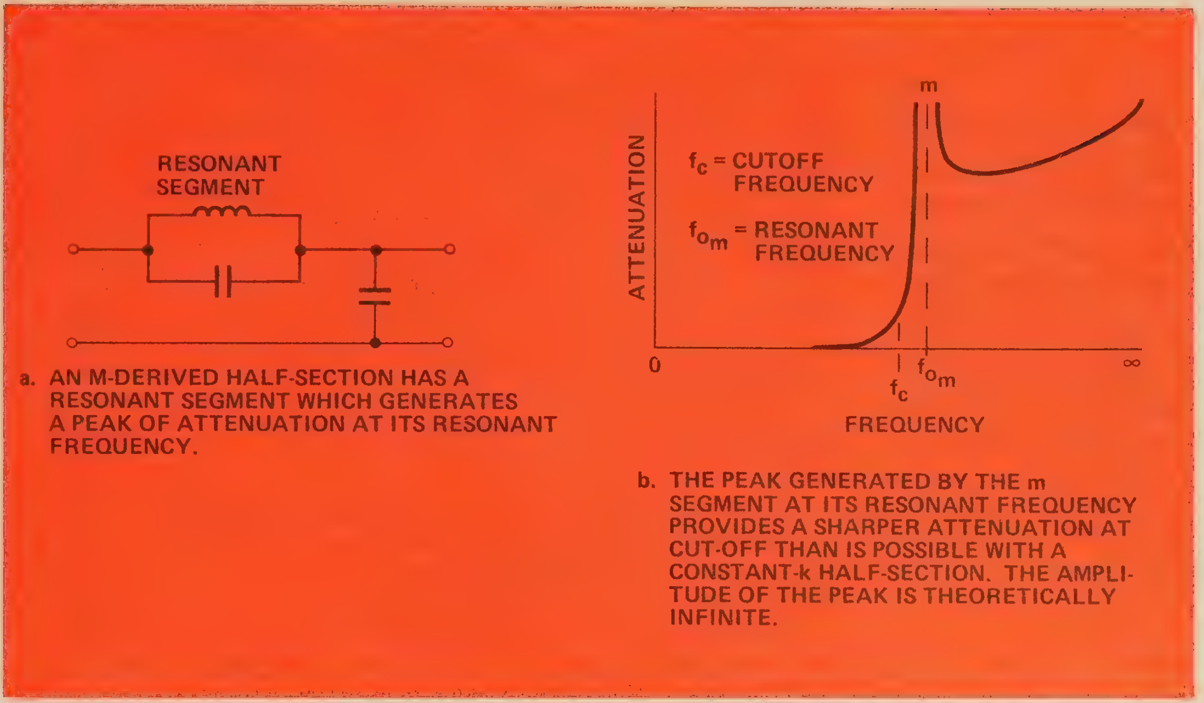


Figure 6. The m-derived filter is the most common image parameter design because of its attenuation peaks. By joining half-sections into a complex filter, several peaks can be established to make optimum use of the stopband with minimum physical components.

pendent upon the attenuation introduced by α. The closer the impedances in the network are matched, the lower the attenuation factor and the greater the power transfer.

To frequencies outside of the passband, Z_i appears as a reactance, which cannot dissipate power. As a result, no power can be transferred and the stopband frequencies are effectively rejected.

Image Parameter Design

The image parameter technique has been in use for so long that its application is fairly well systemized. The first step in designing an image parameter filter is the use of templates to generate a curve, on specially designed graph paper, representing the required image attenuation constant, α, in dB, and the points in the frequency spectrum at which the m-sections will be resonant to produce their characteristic peaks. The m-derived section is much more commonly employed in image parameter filters than constant-k types because of its rapidly rising and frequency-variable attenuation peaks.

The design technique is basically a trial-and-error method in which the templates are moved around until the sum of all the stopband peaks approximates the network attenuation requirements (see Figure 7). Then, because a perfect impedance match is impossible to achieve, slight adjustments – "fudge factors" – are introduced to make the design more realistic. Once the stopband peaks have been established, the value of m which will provide each peak is known. From this, it is possible to calculate the component values that would resonate at the proper frequencies, and thus arrive at an optimum design for the m-sections. Values for the other elements in the filter can be derived from the known network cutoff frequency.

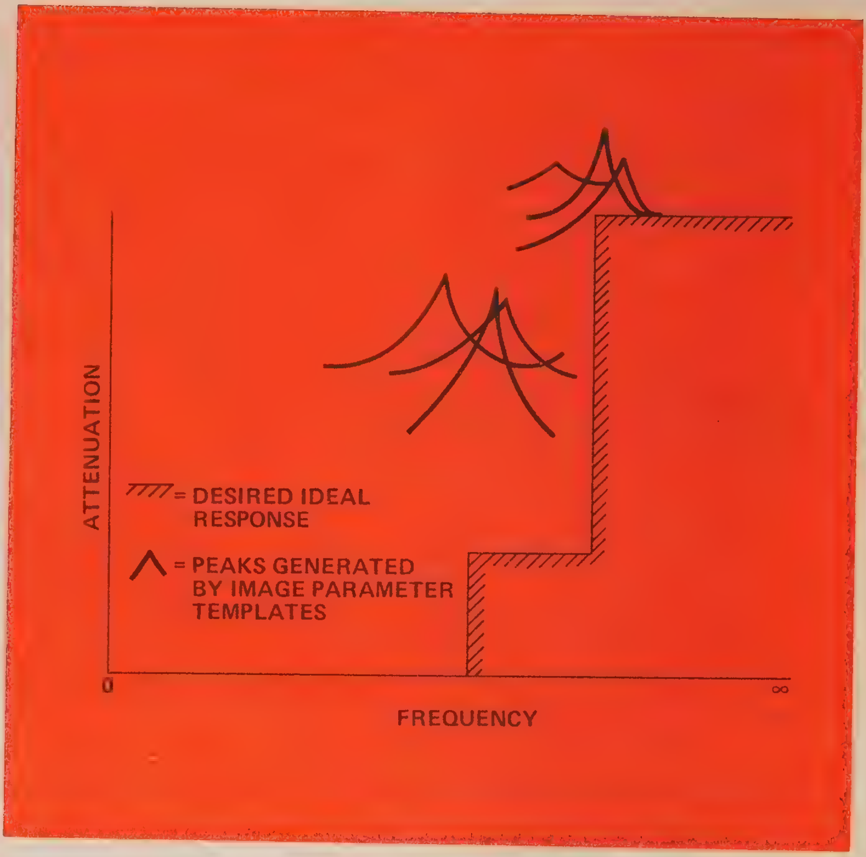


Figure 7. In designing an image parameter filter, attenuation peaks are generated by templates on special graph paper to approximate the desired ideal response, as shown by these approximations for a low-pass filter.

When the image parameter-designed filter is built, it provides a fairly good impedance match to the driving and loading impedances in the passband. When the match is exact, the passband response is at its best – that is, attenuation is minimum. This condition, however, cannot be achieved with any certainty in practice because real network terminations are generally resistive, rather than frequency-dependent impedances. Thus, although an impedance match may be obtained when the network image impedances equal the resistive terminations, the match at other frequencies in the passband is not exact, which can result in badly degraded response. This degradation causes an arbitrary ripple shape to appear in the passband, making it difficult to predict the response of a given filter before it is actually built (see Figure 8). For this reason, image parameter design is today generally limited to constant resistance networks such as all-pass filters. For other applications, modern passive filter design is accomplished with synthesis techniques.

Filter Synthesis

Modern filter design concerns itself with idealized mathematical models which describe filters in terms of complex current and voltage relationships. The elements in the models are defined in terms of a complex frequency variable, s; when the model equations have been solved, mathematical transformations are used to extract component values which would realize the model in physical terms.

The complex frequency variable is expressed as:

$$s = \sigma + j\omega$$

where σ is a real component and $j\omega$ is an imaginary component. It should be noted that the imaginary factor is not physically imaginary; the terms "real" and "imaginary" are mathematical designations for two distinct parts of a

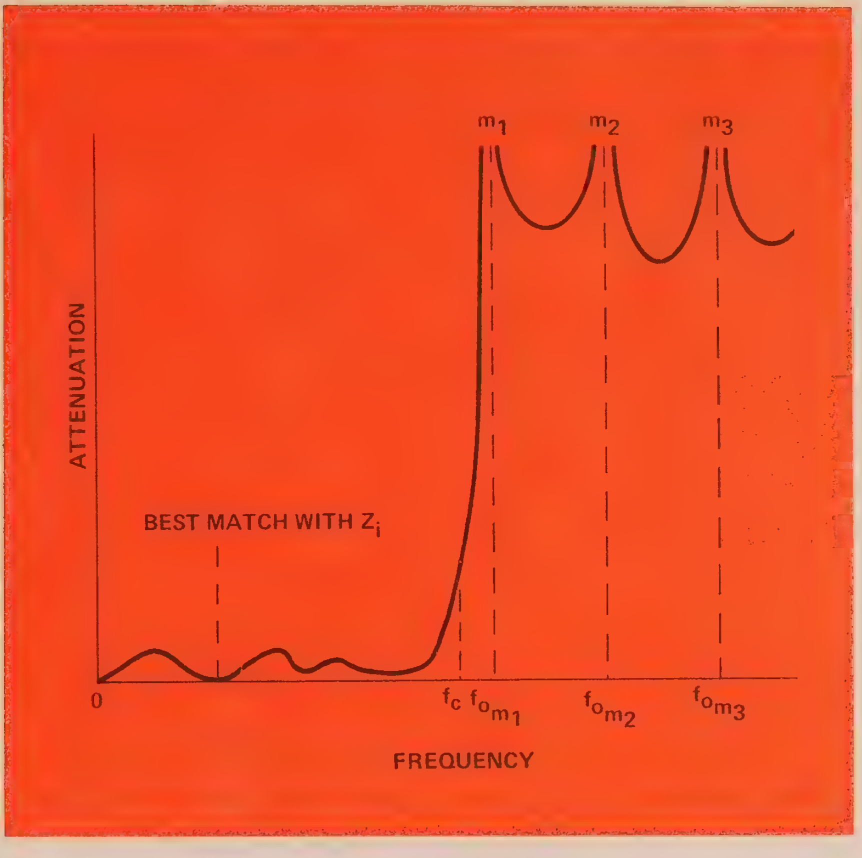


Figure 8. The passband response of an image parameter-designed filter may exhibit an arbitrary ripple shape, as shown in this response curve for an m-derived filter with three m-resonant components.

complex quantity or function, and do not indicate actual existence or nonexistence. A term is considered imaginary when it is related to a strictly mathematical construct, $\sqrt{-1}$, which is represented by the symbol j. Thus, in the expression $s = \sigma + j\omega$, the "real" and "imaginary" factors define two actual aspects of a particular complex frequency.

In relation to frequency, an ideal filter is one which produces no loss of transmitted energy within its passband, yet provides infinite attenuation of stopband frequencies. Such ideals are, of course, unrealizable, given the non-ideal nature of physical components. The problem, therefore, is to achieve a filter design which approximates as closely as possible the ideal for a given application.

Transfer Function

Synthesis filter design begins with an attempt to find a transfer function providing the best approximation of the required filter attenuation and phase properties. The transfer function expresses a mathematical relationship between a filter output quantity and an input quantity; most commonly these quantities are voltages, but they could as easily indicate a current-to-voltage or voltage-to-current ratio. Essentially, the transfer function is a measure of how efficient – or inefficient – a filter is at transferring a quantity at its input to its output port.

In discussing passive filters, it is most convenient to consider the transfer function as an input-to-output ratio; this allows attenuation, rather than gain, to be dealt with.

Represented by the symbol T(s), the transfer function of a passive filter can be expressed in the form:

$$T(s) = \frac{P(s)}{Q(s)}$$

where P(s) and Q(s) are polynomials – algebraic terms consisting of a con-

stant multiplied by variables – representing the input and output quantities, respectively. Through various mathematical manipulations, the P and Q polynomials can be treated in such a way that their roots – those quantities which are multiplied together to form the algebraic terms – take on frequency characteristics. Designating the roots of P as p_1, p_2, etc., and the roots of Q as q_1, q_2, etc., the transfer function can be written:

$$T(s) = \frac{(s-p_1)(s-p_2)\ldots(s-p_p)}{(s-q_1)(s-q_2)\ldots(s-q_q)}$$

where s is the complex frequency variable and the polynomial roots are complex frequencies. Thus, when the value of complex frequency variable s equals the value of any of the roots of P, the numerator becomes zero, resulting in a zero transfer function. For example, if the complex frequency represented by p_1 is the same as that represented by s, then $(s-p_1) = 0$. The product of any number of factors multiplied by 0 is still 0, so $(s-p_1)(s-p_2)\ldots(s-p_p) = 0$. Zero divided by any other number is zero, so where $s = p_1$, the transfer function T(s) is zero, which also is true when s equals any of the other roots of P.

By the same reasoning, when s equals any of the roots of Q, the denominator becomes zero and T(s) has an infinite value (any number can be infinitely divided by zero).

Depending upon what quantities are being related, p_i and q_i (any of the roots of P and Q) represent poles or zeros; if gain is being considered, p_i identifies a "pole of transmission" and q_i identifies a "zero of transmission." When dealing with passive filters, however, it is more convenient to consider attenuation characteristics, since there is no gain in a passive filter; p_i in this case is used to identify a "zero of attenuation" and q_i a "pole of attenuation." This is, in fact, a logical step, since an attenuation pole – a point at which infinite attenuation is presented to a signal – would necessarily produce a zero of transmission. In the following discussion, the poles and zeros of attenuation are being considered. It is thus possible to define a transfer function in terms of its composite poles and zeros, reducing the need to cope with complex polynomial expressions.

Poles and zeros are plotted in what is called the s-plane, where the horizontal axis is σ and the vertical axis is $j\omega$ (see Figure 9); the critical frequencies at which poles appear are theoretically points of infinite attenuation and the zero appearances are critical frequencies exhibiting zero attenuation. The zero critical frequencies are the points at which the filter would oscillate. If the zeros were to appear in the right half of the s-plane, it would indicate that the response of the network grows without bound for any input; in other words, the network would be unstable. A zero appearing on the $j\omega$ axis would also imply that the network oscillates with no input signal. Since both of these events are impossible with passive elements, the zeros are limited to the left half of the s-plane plot.

Synthesis Design

The first step in the synthesis filter design technique is solving what is called the approximation problem.

Since the network being synthesized consists essentially of lumped reactances terminated at each end by a resistor (see Figure 10), the power applied to the network must either be transferred to the load or reflected back to the source, since no power can be dissipated in an ideal reactance. In actual practice, complete power trans-

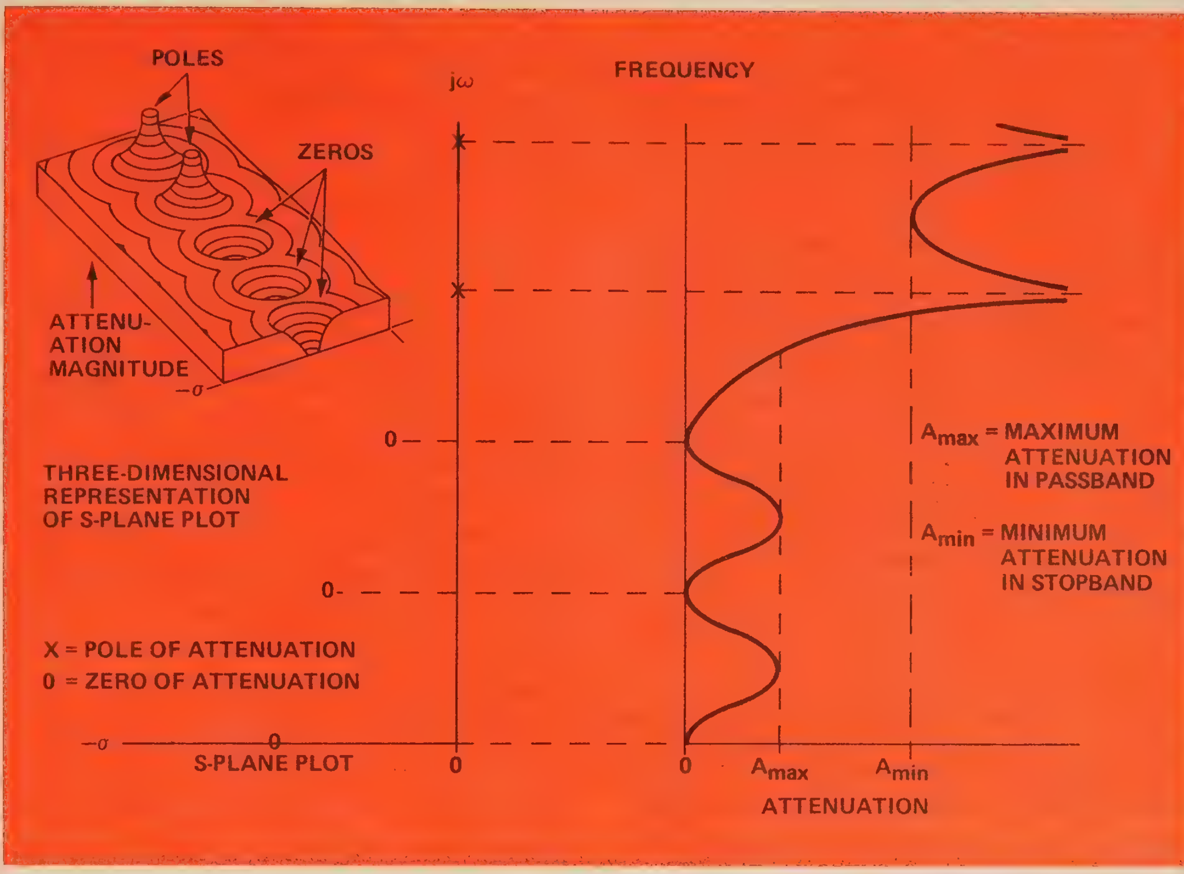


Figure 9. From pole-zero information, it is possible to ascertain the response of a projected filter design, as shown in this plot for a lowpass filter. In practice, filter design is most commonly done with mathematical formulae rather than plots.

fer does not occur, so there is always a reflected component to be considered in the design of a filter; this component is expressed in terms of a reflection coefficient, ρ_i. The reflection coefficient can be defined as the ratio of two polynomials representing maximum possible power (e) and the difference between maximum possible power and power actually delivered to the load (f). Expressed in relation to the complex frequency, these polynomials are related by:

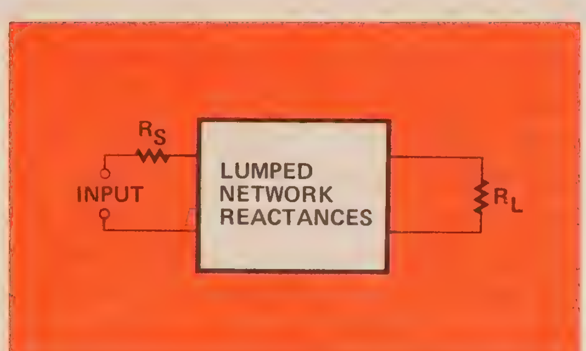


Figure 10. A filter network is essentially a group of reactances terminated by a load (R_L) and source (R_s) resistance. Unlike an image parameter-designed filter, the terminations in synthesis design are not constrained tq match the network impedance characteristic.

$$\rho_i = \frac{f(s)}{e(s)}$$

Under actual design conditions, ρ_i is usually given as a percentage figure: a 100% reflection coefficient means that all of the power is reflected back to the source, and a 0% coefficient means that all of the power is transferred to the load.

The total power in a reactance network is always constant, and can be

determined from the sum of the reflected and transferred powers. This is stated in the Feldtkeller relationship:

$$H(s)H(-s) = 1 + K(s)K(-s)$$

where H is a transmission transfer function having to do with the maximum transfer of power, and K is a characteristic function indicating loss in the network. These functions are defined in terms of the reflection coefficient: H is the ratio of the maximum power available for delivery to the load, e(s), to the power actually delivered, p(s):

$$H(s) = \frac{e(s)}{p(s)}$$

and K is the ratio of maximum deliverable power minus power actually delivered, f(s), to power actually delivered to the load:

$$K(s) = \frac{f(s)}{p(s)}.$$

Thus, the transmission and attenuation functions can be treated as ratios of polynomials capable of generating poles and zeros. Since the poles and zeros are contained within the H and K terms, it is not necessary to create an s-plane plot to deal with them; they can be more easily handled in this form by mathematical manipulation.

When a filter is specified, it is usually in terms of its loss or its phase/delay response. If the loss response is known, a value for K can be determined and from this, H can be derived; K can likewise be obtained from H when the phase/delay response is specified. Once values of H and K are known, both sets of polynomials can be factored into odd and even parts, and the parts combined to give required input impedance functions. For example, dividing the difference between the even parts of H and K by the sum of the odd parts gives a value for the input impedance of port 1 with a far-end open circuit (Z_{11}):

$$\frac{H_{even} - K_{even}}{H_{odd} + K_{odd}} = Z_{11}.$$

From the impedance functions, component values can be calculated. In practice, most of these design steps are now done by computer calculation, greatly increasing the complexity of the filters which can be designed, and providing the special mathematical handling required to conserve accuracy in the design of large filters.

Design Realization

Once a filter has been designed, it must be constructed – realized – using physical components. During the design process, these components are assumed to be resistors, capacitors and inductors; in reality, the resistive element is quite often present in distributed form only, such as the resistance in an inductor's windings, so that the realization of an RLC filter network is generally composed of only inductors and capacitors. Figure 11 shows typical realizations for four of the most common filter types: the low-pass, which rejects signals above a specific cut-off frequency, the high-pass, which rejects signals below cut-off, the band-pass, which rejects signals on either side of a selected passband, and the band-stop, which passes signals on either side of a selected stopband. Among the myriad uses of such LC filters are receiver input preselection, suppression of unwanted sidebands and harmonics, impedance matching, and multiplexing.

The realization of a filter design is not required to consist of inductors and capacitors. Depending upon overall circuit applications, portions of the network may be replaced by crystals,

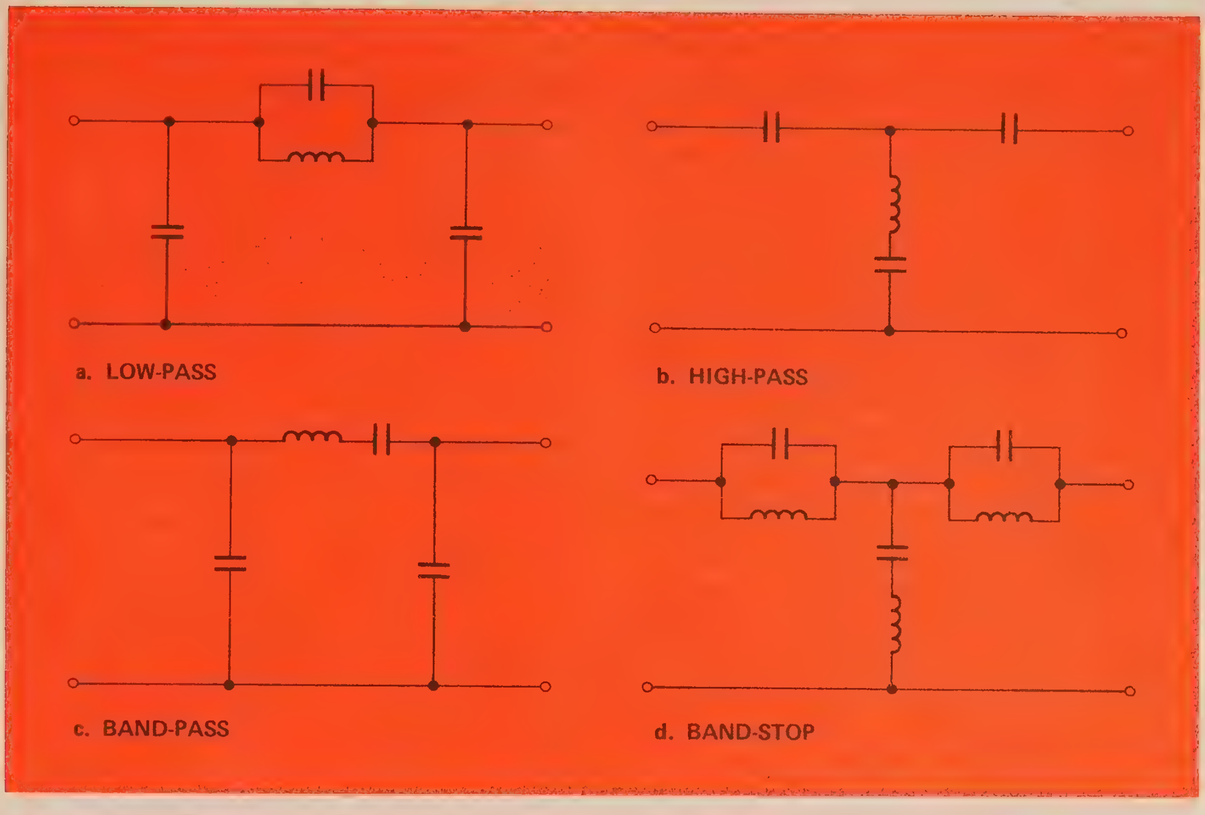


Figure 11. The basic filter types can be realized as strictly LC structures, as these typical networks illustrate.

mechanical resonators, and operational amplifiers (for active filters), or the entire network may consist of a waveguide cavity or a space in a strip transmission line.

Crystal Filters

Certain crystalline materials – notably quartz – exhibit a piezoelectric property, in that they can be set in mechanical resonance by an electrical field; the frequency at which this resonance occurs is a function of the crystal's size and the manner in which it is cut. This transducing property, which allows translation of electrical signals into mechanical energy and vice versa, makes it possible to realize a filter with crystal resonators – plates of piezoelectric material with a metal electrode on each side – rather than electrical components, since the frequency at which the crystal resonates is effectively passed while adjacent frequencies are blocked.

In designing a crystal filter, the image parameter and/or synthesis techniques are first used to obtain component values for an equivalent LC filter; information on crystal size and shape, electrode dimensions, and electrode placement is then used to transform the LC network into a crystal resonator. Several such resonators can be coupled together to form a filter with the response characteristics required by a given application.

Crystal filters are widely used in the telecommunications industry to provide channel separation in multiplex systems because of their low cost and size reduction compared to LC networks. These advantages are obvious in those structures which combine several elements into one physical unit, such as GTE Lenkurt's polylithic filter.

Microwave Filters

The uses of filters in the microwave frequency range are the same as those

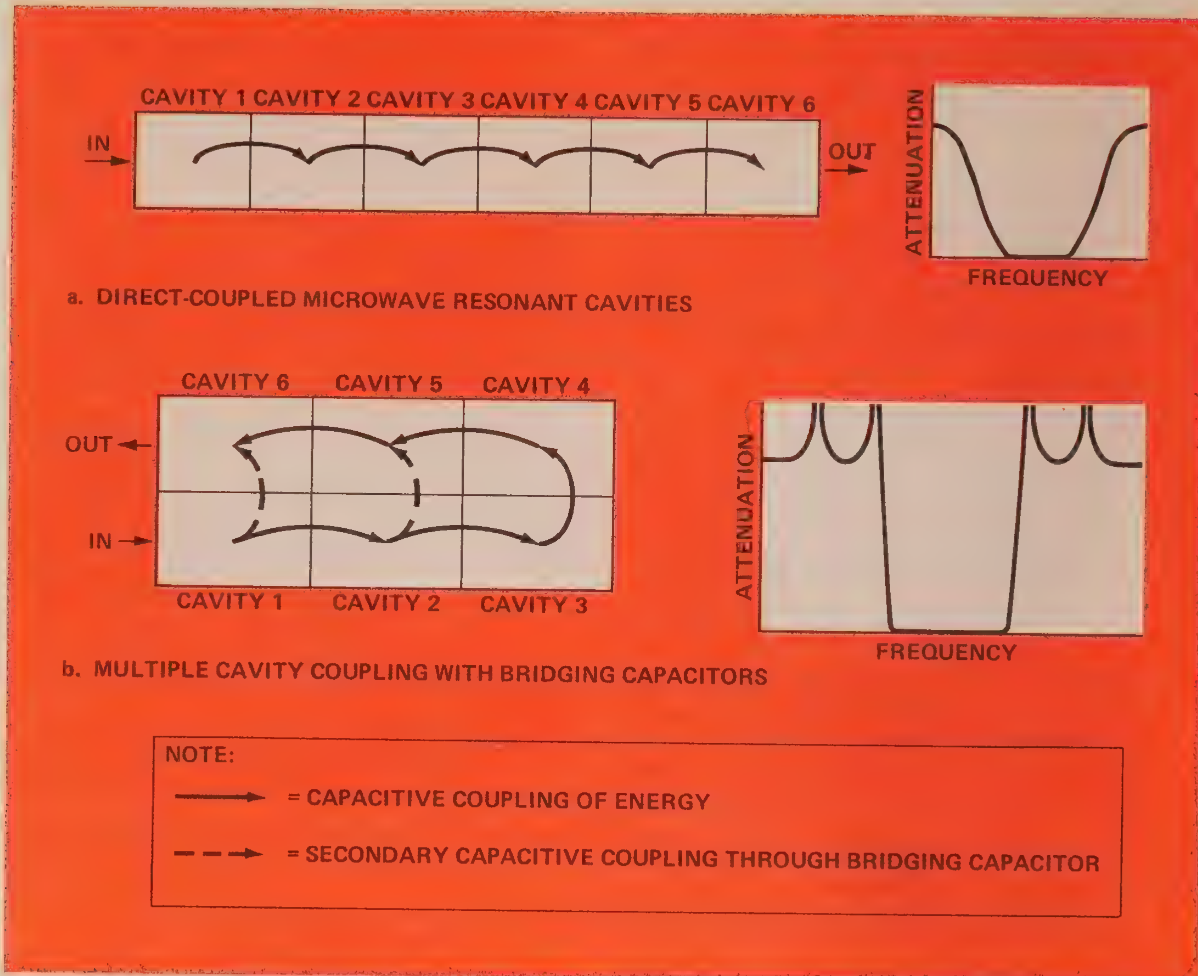


Figure 12. In an experimental approach to multi-cavity filters, bridging capacitors allow cancellation of portions of the microwave energy, producing attenuation peaks which improve the overall filter performance.

at lower frequencies: they can be used to select, reject, and channel electromagnetic energy of different frequencies, and provide for maximum transfer of that energy from one point to another.

In the design of microwave filters, the lumped inductive, capacitive, and resistive components of lower-frequency networks are replaced by the characteristics of such typical microwave circuit elements as waveguides, coaxial and strip transmission lines, and resonant cavities. The nature of these microwave elements is such, however, that under certain conditions their frequency behavior can be made to approximate that of lumped components. Because of this, the most common approach to microwave filter design is the derivation of structures from equivalent lumped-element filters generated by image parameter or synthesis techniques. Indeed, if a lumped-element filter is designed to have the same frequency response characteristics as a microwave filter, it is possible to replace almost every lumped element with an equivalent microwave element.

Although derived from them, microwave filter realizations do not physically resemble lower-frequency LCR networks at all. The elements seen in microwave circuits consist of such devices as stubs — short- or open-circuited sections of transmission line used as reactive components —

and resonant cavities – regions of dielectric surrounded by conductive walls. For example, low- and band-pass filters used at lower microwave frequencies can be realized by open- and short-circuited sections of coaxial line. At higher frequencies, resonant cavities can be coupled together to produce required responses. In one such application, waveguide sections are constructed with cavities built into them; to allow for adjustment of each cavity's resonant frequency, screws are often extended through the waveguide wall to vary the capacitance of the cavity. Variations of this technique are currently under development which will produce a filter frequency response whose stopband contains the attenuation peaks – and the advantages these provide – found in m-derived sections at lower frequencies. One of these new techniques involves coupling resonant cavities not only in cascade, but also in parallel (see Figure 12), allowing cancellation of energy at different points in the filter.

This discussion has been concerned with the design of passive filters – filters with no source of energy within the network – according to image parameter and synthesis techniques. In Part 2, the design of active filters will be discussed.

BIBLIOGRAPHY

1. Harvey, A.F. *Microwave Engineering.* New York: Academic Press, 1963.

2. Lampitt, R.A. *Network Analysis For Telecommunications and Electronics.* London: Iliffe Books Ltd., 1965.

3. Orchard, H.J. and G.C. Temes. "Filter Design Using Transformed Variables," *IEEE Transactions on Circuit Theory,* Vol. CT-15, No. 4 (December 1968), 385 - 408.

4. Sheahan, D.F. and R.A. Johnson. "Crystal and Mechanical Filters," *IEEE Transactions on Circuits and Systems,* Vol. CAS-22, No. 2 (February 1975), 69 - 89.

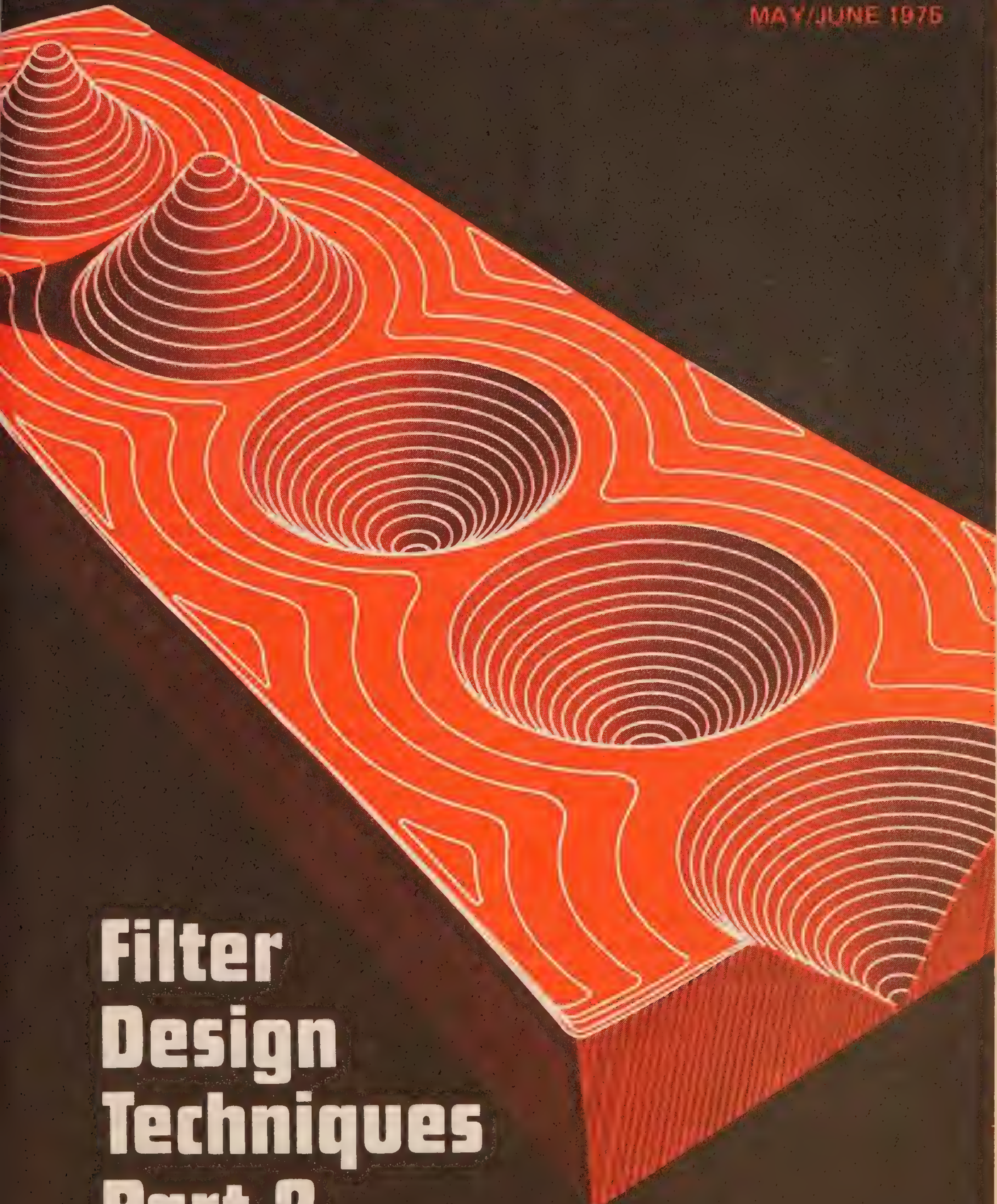
5. Skilling, H.H. *Electrical Transmission Lines.* New York: McGraw-Hill Book Co., Inc., 1951.

6. Van Valkenburg, M.E. *Network Analysis.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1964.

7. Zverev, A.I. *Handbook of Filter Synthesis.* New York: John Wiley and Sons, Inc., 1967.

GTE LENKURT DEMODULATOR

MAY/JUNE 1975



Filter Design Techniques Part 2

Progress in the miniaturization of electronic circuitry has necessitated new approaches to the selective processing of electromagnetic waves. Among the most effective are those which make use of active elements in RC networks.

The April, 1975, issue of the *Demodulator* began a discussion of the techniques used in designing electromagnetic wave filters. It treated them in terms of their resonance characteristics and attenuation functions, restricting itself to filter networks containing only passive components. In this issue, the discussion is continued, concentrating on the design of active filters having lumped resistive-capacitive (RC) components; this is one of four basic types of active filter, the other types being: active filters with distributed RC components; N-path filters, in which network parameters are time-variable; and digital filters, in which digital processing structures are used as signal-wave filters.

RC Networks

In the design of passive filter networks, it is assumed that the physical construction – the realization – of the design will consist of inductors, capacitors, and resistors connected in an RLC network. Since the functional resistive element may only be present in a distributed form, such as the resistance in an inductor's windings, rather than as a lumped component, it is a common practice to consider only the reactive, or LC, filter elements. In many cases, however, the physical structure of inductors makes the realization of such passive RLC and LC networks extremely difficult; for example, the bulk of a low-frequency inductor makes it incompatible with many of the modern integrated and printed circuit techniques. As an alternative, filters can be designed without an inductive element; that is, as strictly RC networks (see Figure 1). Such networks have the advantage of small physical dimensions, low cost, and relative insensitivity to interference from external electrical forces.

Without an inductive element, however, the properties of the RC network cause its frequency selectivity to be very poor and limit its range of application. In addition, it requires a large number of components to produce a response equivalent to that of an LC network, thus negating its size advantage. These deficiencies can be overcome through the use of one or more active devices, which are essentially amplifiers, as elements of the filter network.

An ideal active RC filter network can be most generally represented as an interconnection of passive resistive and capacitive elements, with active devices serving to compensate for the absence of inductors; the function of such a network is to selectively process

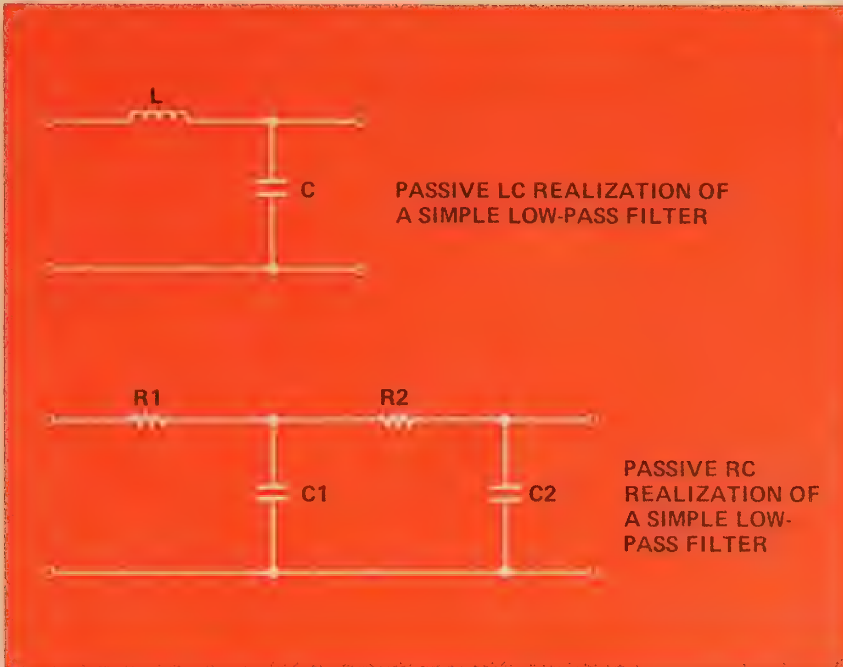


Figure 1. Passive filters can be realized as either inductive-capacitive or resistive-capacitive networks.

signals according to their frequency composition.

Techniques for designing active RC filters fall generally into two broad categories: simulation, which normally requires that an active circuit replace and perform the function of an inductor or other energy-storage element, and biquadratic filtering, in which the transfer function required in a given application is broken down into biquadratic factors – mathematical components of the transfer function – and filter sections realizing these factors connected in cascade.

Active Filter Simulation

The design of an active RC filter by simulation involves two basic steps: (1) designing a passive LC filter with either the image parameter or synthesis method (see the April, 1975, *Demodulator* for a discussion of these techniques), and (2) replacing each inductor in the passive LC filter with an active RC circuit capable of simulating the response of the replaced inductance. A variation of the second step results in the replacement of the inductors by resistors, resistors by capacitors, and capacitors by active circuits called frequency-dependent negative resistances (FDNR's).

The simulation approach allows use to be made of the well-established passive filter design techniques, which simplifies the process, and also takes advantage of the LC filter's inherent frequency stability and normally low sensitivity to variations in component value. The latter is quite often the deciding factor in determining what type of active circuit will be utilized in a given filter realization.

The problem of stability does not occur in passive filter networks because they contain no internal power sources. An active filter, however, may break into oscillation when any of its parameters, such as the values of passive components or the gain of active devices, change. The network transfer function of an active filter is also sensitive to changes in these parameters; even when stability is not affected, such variations may produce filter behavior totally different from the design requirements. The active circuits used in the simulation technique are thus evaluated not only in

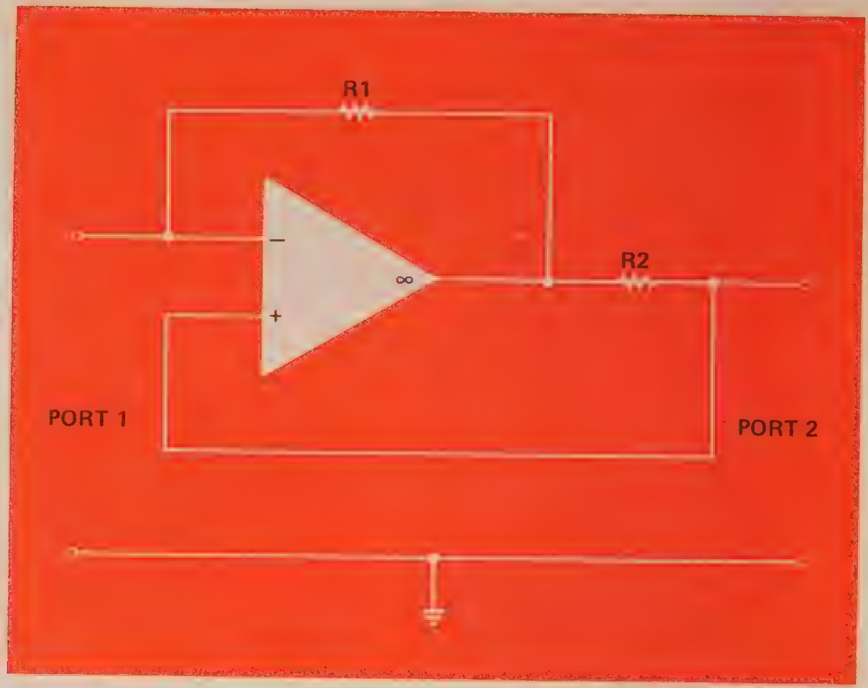


Figure 2. Realization of a current-inverting negative impedance converter (INIC) using an operational amplifier.

terms of their simulation properties, but also according to their sensitivity to parameter changes.

Once the LC filter is designed, it is necessary to decide upon the most suitable inductor-simulating circuit. Because an inductor is essentially an energy-storing device, any simulating circuit must also be able to store energy; the only other device with this capability is the capacitor, so an inductance simulator must of necessity contain some capacitive element.

Since the initial passive design procedure yields a complete filter configuration, active RC filter simulation has concentrated on the realization of an effective inductance simulation circuit. Most of the resulting designs have fallen into the positive- and negative-impedance converter or inverter categories, and have included such circuits as the negative impedance converter and the gyrator. In the last few years, all of these design approaches have been related and re-defined in terms of a "generalized impedance converter" (GIC), whose properties encompass the characteristics of the converter/inverter circuits.

Negative Impedance Converter

An impedance converter is a two-port network whose function is to change an impedance characteristic in some manner. Ideally, when the impedance converter is terminated at one port by an impedance, Z, the input impedance at the other port is directly proportional to Z for all frequencies; the proportionality in a network containing only an amplifier and resistors is determined by a conversion factor, or impedance transformation function, K. If an energy-storing device such as a capacitor is placed in the network, as is often the case, a complex frequency variable, s, appears in the relationship, making the conversion factor K(s). A capacitor serving as a terminating impedance does not introduce the frequency variable, so in this discussion, which deals with the simplest realizations, the (s) notation is not used.

Negative impedance is a characteristic of some circuits and components which causes a decrease in current for an increase in voltage, and vice versa; this property has been recognized for some time, and is exhibited by such devices as tunnel diodes and, under

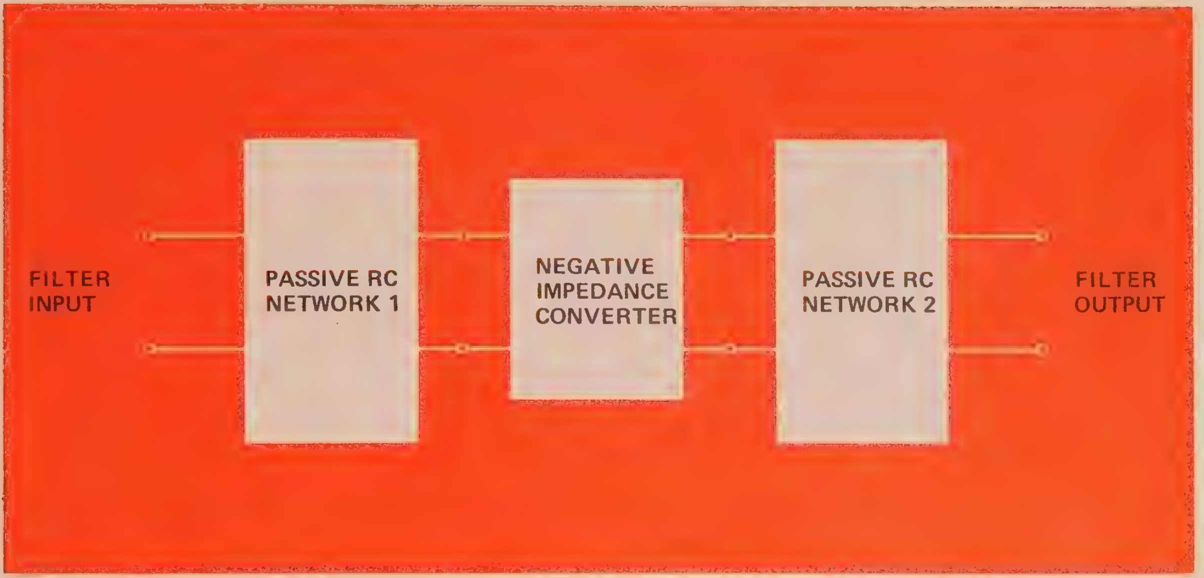


Figure 3. A negative impedance converter is typically placed between two passive RC networks to form a filter section.

certain conditions, vacuum tubes. A negative impedance converter (NIC) is a two-port network which converts an impedance terminating one port into its negative at the other port. For example, if port 2 were terminated by impedance Z_L, the input impedance, Z_{IN}, at port 1 would be related to it by

$$Z_{IN} = -KZ_L$$

where K is the network conversion factor.

A realization of an NIC is typically composed of one operational amplifier (op amp) and associated resistors (see Figure 2). The circuit is placed between two passive RC networks to form a filter section (see Figure 3), so that the negative impedance of network 2 created by the NIC can interact with the positive impedance of network 1. This interaction produces the desired filter response characteristic.

The op amp is generally preferred for active filter circuit realization because of its operating characteristics, its size reduction, the ease with which it can be incorporated into integrated circuits, and its ready availability as an inexpensive component. Theoretically, an op amp has infinite input impedance, zero output impedance, and infinite gain; these characteristics, along with the ability to produce a single output proportional to two inputs (see Figure 4), make the op amp ideal for active filter applications.

Gyrator

Impedance inversion is a property allowing the input impedance of a two-port network to be inversely proportional to its terminating impedance (see Figure 5). As with an impedance converter, the proportion is such that when one port is terminated by Z_L, the input impedance, Z_{IN}, at the other port is related by a conversion factor, K. The relationship in an impedance inverter is expressed as:

$$Z_{IN} = K\frac{1}{Z_L}$$

A gyrator is an impedance inverter in which K is a positive real constant, giving the gyrator the ability to transform a terminating capacitive reactance into an inductive reactance at its

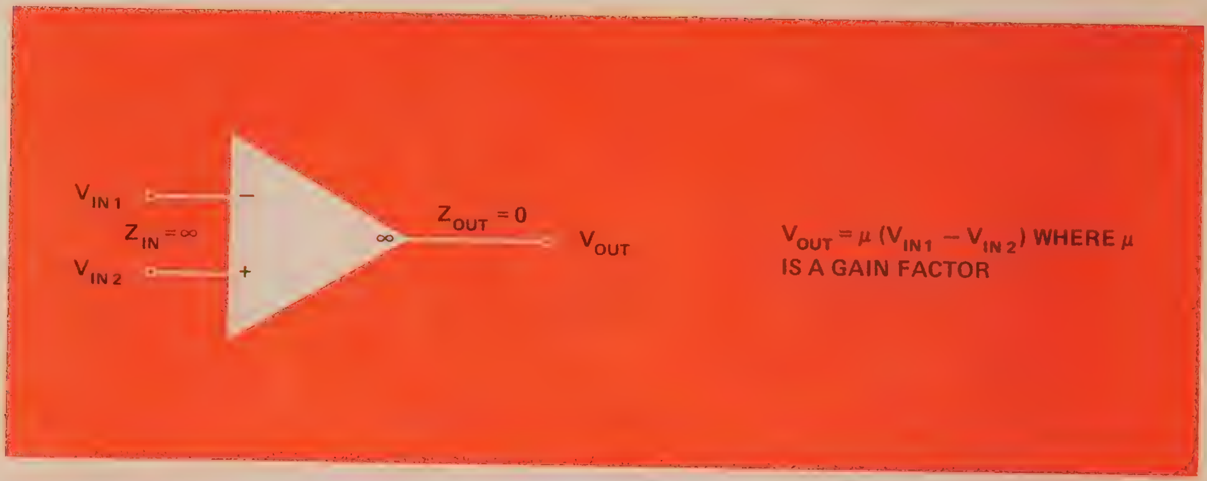


Figure 4. An operational amplifier is ideal for active filter realizations because of its impedance and gain characteristics.

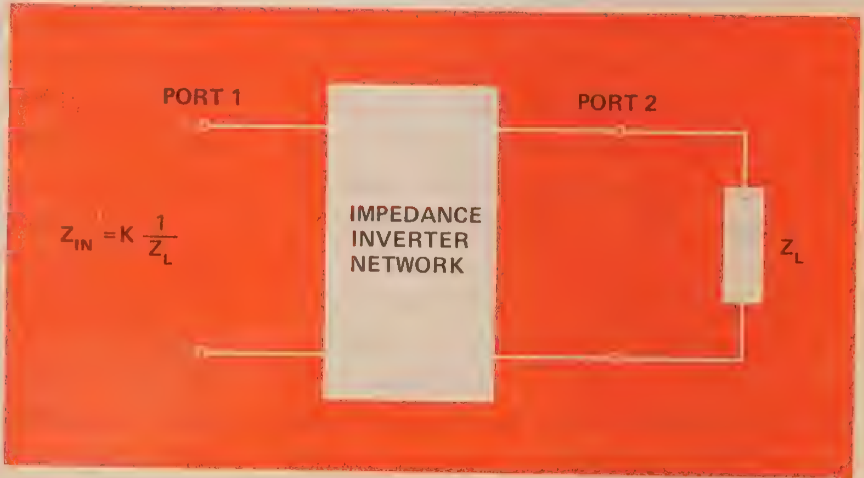
input. Typically, a gyrator is realized as two op amps connected as parallel current-to-voltage converters (see Figure 6).

Generalized Impedance Converter

The properties of the impedance converter — both the positive and negative types — and of the impedance inverter — both the positive type, of which the gyrator is one realization, and the negative type — have been redefined and incorporated into the more universal GIC. This structure is defined as a two-port network whose input impedance is the product of its terminating impedance and some internal network function, which is actually the K factor. According to this definition, if the GIC were to be terminated at port 2 by Z_L, Z_{IN} at port 1 would be equal to KZ_L; if port 1 were terminated by Z_L and port 2 used as the input point, Z_{IN} would be defined by the term Z_L/K. The relationships are the same as those developed for the converter/inverter circuits, but they are here embodied in one network.

The realization of a GIC can be accomplished in several ways, one of which is shown in Figure 7. The various impedance elements can be physically introduced as capacitors or resistors, depending upon the simulation characteristic desired. In Figure 7, for example, making Z2 a capacitor and all of the other elements resistors

Figure 5. An input impedance inversely proportional to a terminating impedance is generated by an impedance inverter.



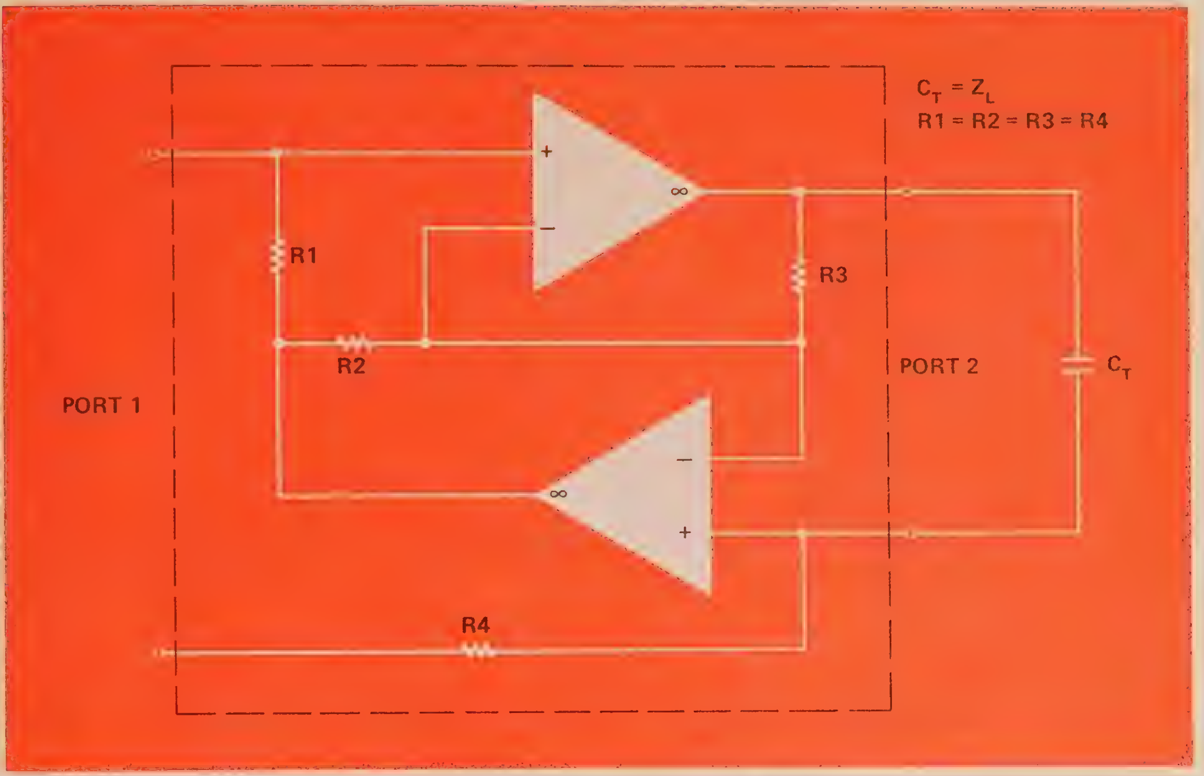


Figure 6. A gyrator can be realized as two current-to-voltage converters connected in parallel to form an impedance inverter network. The capacitive reactance terminating port 2 is made to appear inductive at port 1.

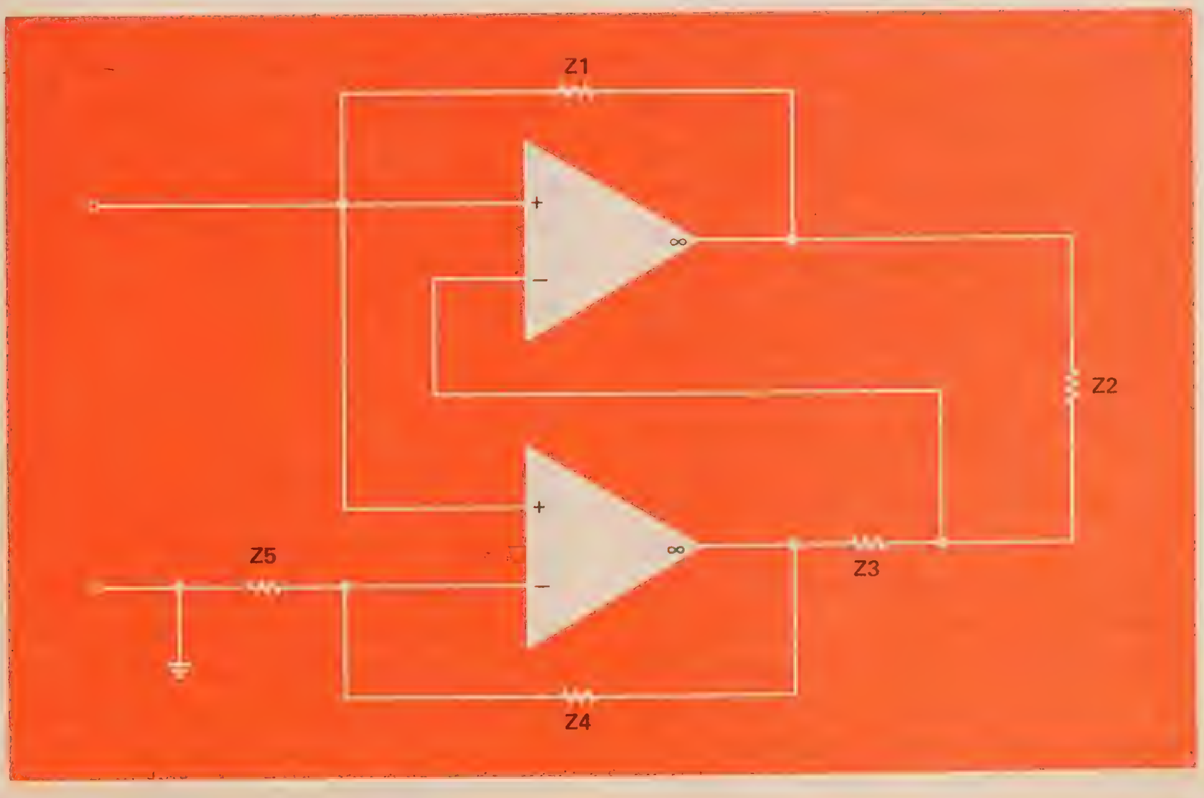


Figure 7. A generalized impedance converter (GIC) is a two-op amp network whose impedances (Z1–Z5) can be realized as resistors or capacitors, depending upon desired response.

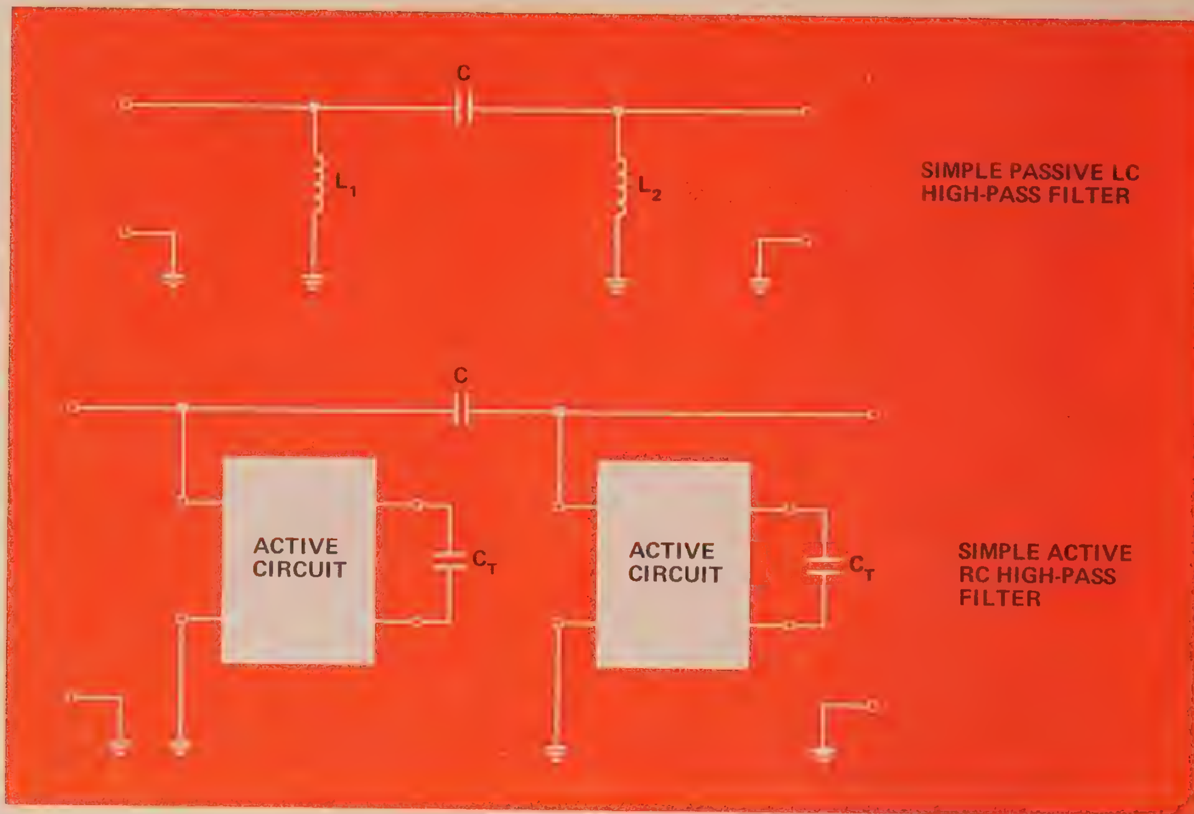


Figure 8. Active circuits commonly appear in filters as two-terminal components. The characteristics of the active circuits, whatever their realized structures, cause their terminating capacitors, C_T, to produce inductor-like behavior.

would result in a configuration equivalent to a gyrator.

Active circuits are most generally incorporated into filters as two-terminal devices (see Figure 8) to introduce inductance characteristics. In some cases, the network is used as a four-terminal device, in which the signal to be filtered is put in at one port and passed out at the other; this technique is shown in Figure 3. Use as a four-terminal device, however, complicates the filter design process because the properties of four, rather than two, terminals must be considered. For this reason, one of the ports is usually terminated and the two terminals at the other port used to connect the network with the other filter elements.

FDNR

A special application of the GIC results in an active circuit called a frequency-dependent negative resistance (FDNR). This is a two-port network which, when terminated at one port by a resistor or capacitor, produces unique impedance characteristics at the other port. The input impedance of an FDNR, or D, network is given by

$$Z_{IN} = -\frac{1}{s^2 D}$$

where s is a complex frequency variable and D is in units of farads squared. From this expression, it can be seen that the impedance characteristic of an FDNR is, indeed, negative and dependent upon the frequency of the signal applied to the simulated filter. Such a circuit can be realized with an appropriately structured GIC, as shown in Figure 9.

The properties of an active filter containing an FDNR are determined

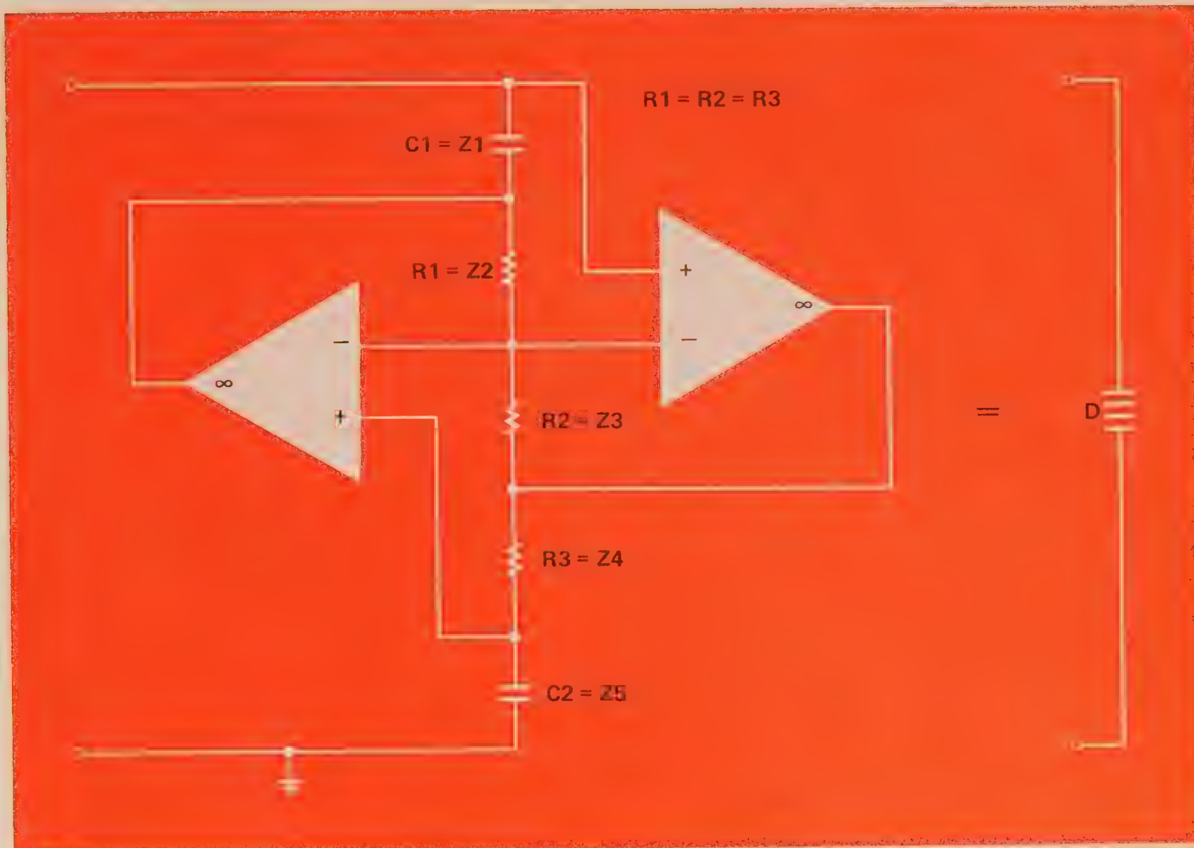


Figure 9. A frequency-dependent negative resistance can be realized as a GIC. The schematic symbol resembles that of a capacitor, but consists of four parallel lines.

through the use of an impedance transformation function, 1/s, in which s is once again the complex frequency variable. Multiplying every impedance in the initially designed LC filter by 1/s allows inductors to be replaced by resistors, resistors to become capacitors, and capacitors to become FDNR's (see Figure 10).

The attempt to create a suitable inductor-simulator for use in active RC filters has produced a variety of circuits, all of which generally depend upon the operational amplifier parameters. An alternate approach to the design of active RC filters is the "biquadratic decomposition" technique, in which the op amp also plays an important role.

Active RC Filter Synthesis

All filters, whether passive or active, can be described in terms of a transfer function, T(s), which is an indication of the efficiency – or inefficiency – of a filter in transferring a quantity at its input to its output; that is, T(s) is a filter transmission characteristic. In passive filters, this function is generally expressed as a ratio of input to output quantity (see April, 1975, *Demodulator*) to allow consideration of attenuation behavior. Active filters, however, contain devices which are essentially amplifiers, so it is more convenient to define their transfer functions as ratios of output to input quantities, or as gain factors. It is also normal practice to consider the quantities to be the input and output voltages, so that T(s) becomes a voltage transfer function.

The problem in active filter synthesis is to arrive at a filter design which, when realized with physical components, produces the transmission char-

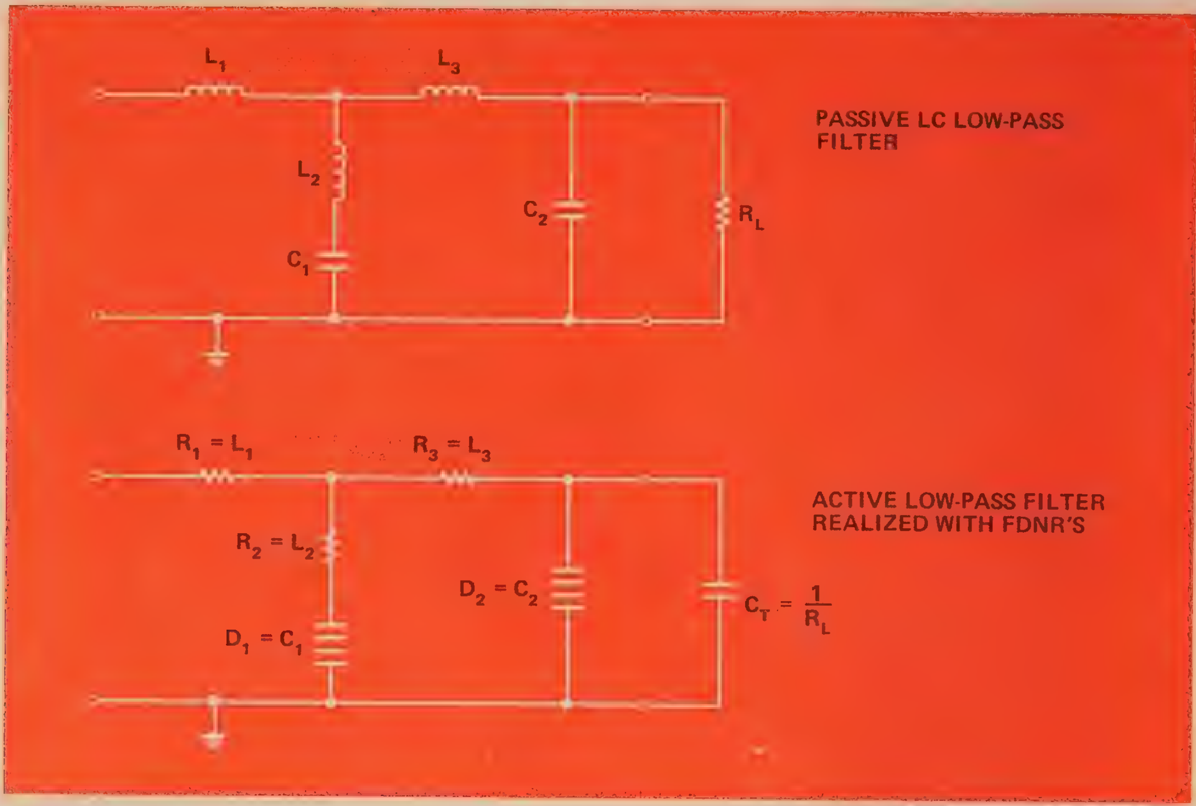


Figure 10. Introducing FDNR's into a filter allows all of the component responses to be simulated by other elements. Simulation techniques such as this result in structures which are very similar in schematic appearance to the passive filters they replace.

acteristic identified by the transfer function.

One approach to this problem requires that T(s) be used to determine an overall filter configuration, just as was done in passive filter synthesis. Another approach, which produces filters whose response characteristics are easier to design, is through realization with biquadratic filter sections.

Biquadratic Filter Realization

As in passive filter synthesis, the transfer function of a given active filter can be expressed as a ratio in the form:

$$T(s) = \frac{N(s)}{D(s)}$$

where N(s) and D(s) are polynomials representing the network output and input quantities, respectively.

Regardless of how complex the polynomial used to identify a quantity, it is possible to break the expression down into second degree, or quadratic, factors of the general mathematical form $P(x) = ax^2 + bx + c$. Performing this operation on both N(s) and D(s) produces a number of quadratic factors in the numerator and a number in the denominator. Each numerator factor is associated with a denominator factor, effectively dividing the filter T(s) into sections, each with its own transfer function t(s), of the form

$$t(s) = \frac{n_2s^2+n_1s+n_0}{d_2s^2+d_1s+d_0}$$

These biquadratic functions — so called because they are defined by two quadratic expressions — are mathemat-

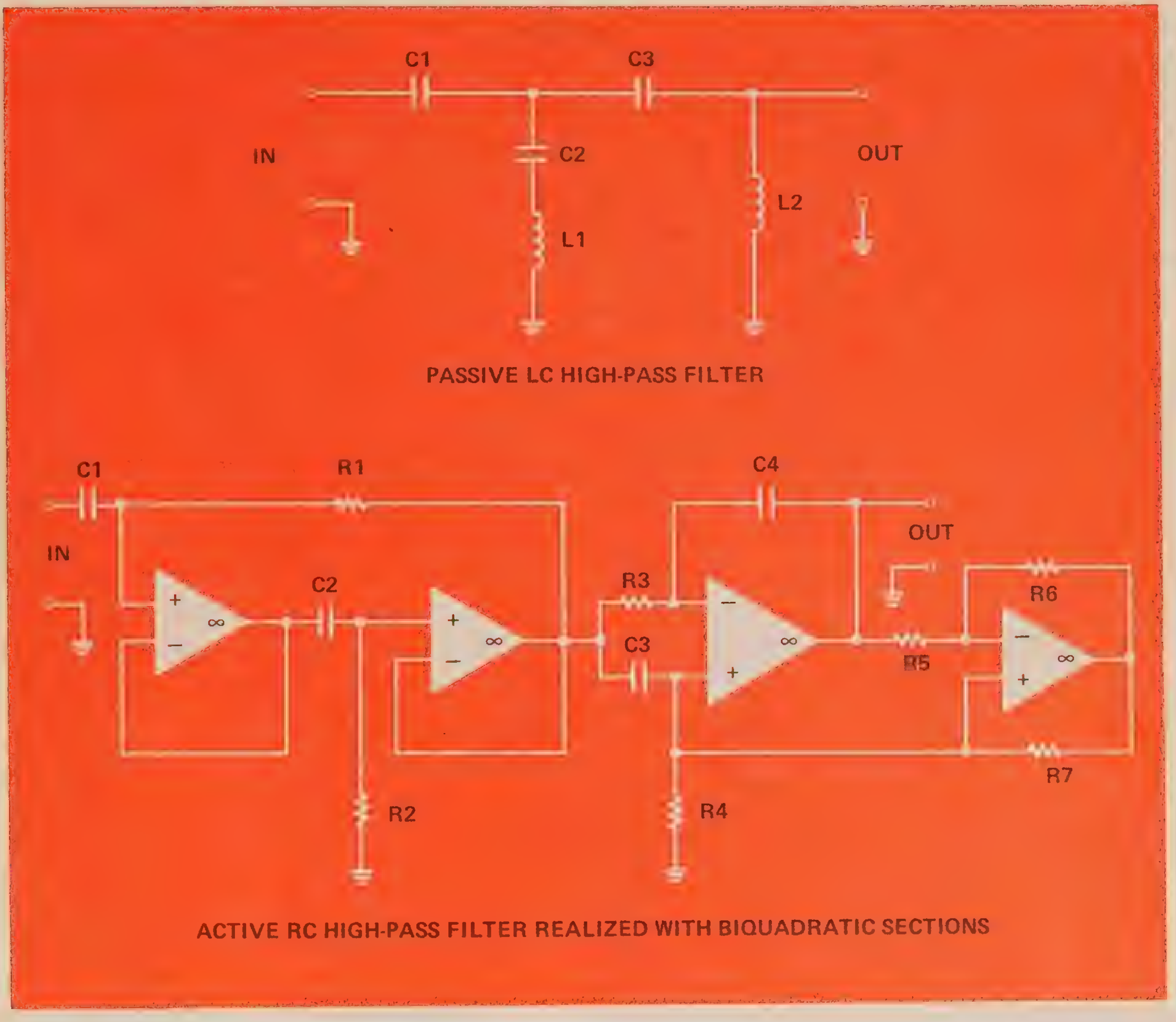


Figure 11. A filter structure can be realized with cascaded biquadratic sections, in which case it may bear very little resemblance to any of the simulation-designed filters.

ically manipulated to produce active filter sections realizing their individual t(s). Connecting the biquadratic sections in cascade – the output of one section becoming the input of the next – produces a complete filter with the desired transmission characteristics.

The physical realization of a biquadratic filter section varies according to a given application's requirements. When these needs have been considered, a decision is reached as to which of the many possible circuits would be most effective. A common biquadratic realization contains a minimum of one and a maximum of four operational amplifiers. The output of the section is taken at the output terminal of one of the op amps, taking advantage of the low impedance at that point to reduce the interaction between cascaded sections (see Figure 11).

When the type of circuit has been selected, it must be determined if the requirements can be met solely by component selection, or if the circuit has to be tuned.

Tuning

Since the amplifiers used in a circuit realization are not perfect, the actual transfer functions produced are likewise imperfect. The best method

for dealing with this imperfection is by tuning the individual biquadratic sections; in this way, resistor and capacitor tolerances can also be compensated for.

Essentially, the tuning procedure consists of making resistance, amplitude and phase measurements on a filter structure, comparing these to the desired quantities, and changing resistance values to make the two values as close as possible. This process can be automated in integrated circuit fabrication by interconnecting a computer with measuring gear and a laser trimmer.

At GTE Lenkurt, the biquadratic filter realization has been used to create the filters used in the 262A Data Set, which processes data at 4800 b/s. At this bit rate, filtering must be extremely precise to avoid pulse distortion; the active biquadratic filter, specifically tuned to match the characteristics of each data set, provides this precision.

Active filters are essential elements in modern electronics applications because of their great savings in size and weight, and their compatibility with integrated circuit techniques. Despite some inherent sensitivity problems, active filters are also valuable because they can not only realize all of the network functions of passive filters, but can also realize transfer characteristics totally unattainable with strictly passive networks.

BIBLIOGRAPHY

1. Bruton, L.T. and D. Treleaven. "Active Filter Design Using Generalized Impedance Converters," *EDN*, Vol. 18, No. 3 (February 5, 1973), 68 – 75.

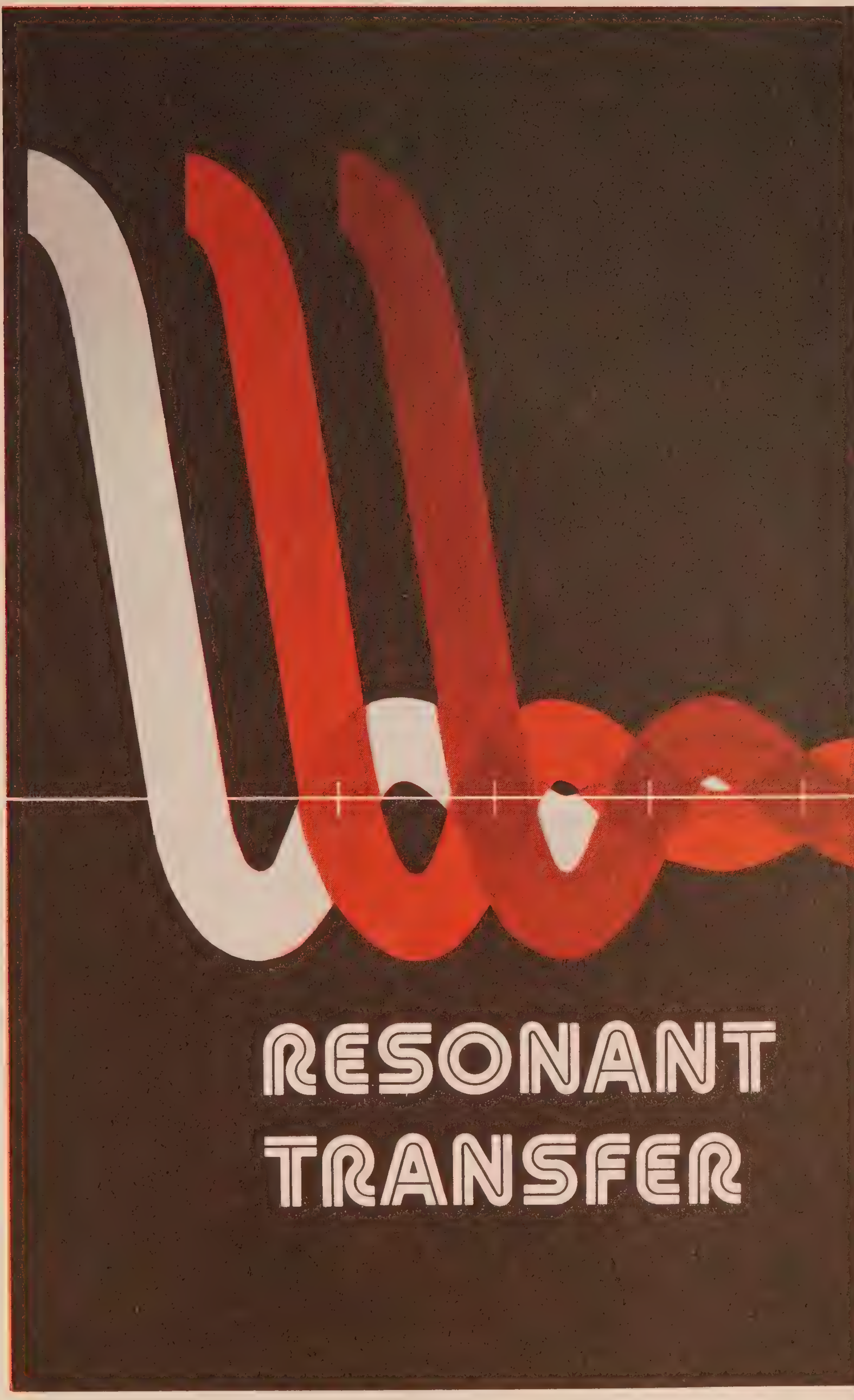
2. Hilberman, D., *et. al.* "A Review of Active Filter Terminology, Technology and Design," paper in preparation for IEEE Circuits and Systems Society. Draft: March 1975.

3. Huelsman, L.P., ed. *Active Filters: Lumped, Distributed, Integrated, Digital, and Parametric.* New York, N.Y.: McGraw-Hill Book Co., 1970.

4. Mitra, S.K. *Analysis and Synthesis of Linear Active Networks.* New York, N.Y.: John Wiley & Sons, Inc., 1969.

5. Schmidt, C. "Active Filters (Technical Memorandum No. 62/73/476)". San Carlos, California: GTE Lenkurt, Inc., 1973.

6. Temes, G.C. and S.K. Mitra. *Modern Filter Theory and Design.* New York, N.Y.: John Wiley & Sons, Inc., 1973.


RESONANT
TRANSFER

Resonant transfer is a novel technique for transferring energy between filters, with minimum energy loss. Use of this technique allows the design of zero-loss amplitude modulators, time-division switching systems, electronic 2/4-wire hybrids which do not require balance networks, simple data transmission filters, and unique FDM carrier systems.

The basic arrangement for the resonant transfer of energy between filters is shown in Figure 1. Generally, when the concept of resonant transfer is applied, it is presumed that energy from a source V is to be applied to an LC filter, FL-A, and is subsequently to be transferred without loss to a second filter, FL-B. The physical arrangement of the filters is such that shunt capacitances appear at the switched ports 2 and 2′, respectively. The two filters are interconnected by a very small inductor L_R, which is in series with an analog gate G. It is convenient for the purpose of explanation to visualize the existence of a series inductor (L_f and $L_{f'}$) within each filter. This inductor is not required for circuit operation but frequently is present as an overall characteristic of the filter and makes the discussion of resonant transfer easier to understand.

Operation of the circuit takes place during two distinct time intervals. One interval, T_s, is comparatively long, on the order of 125 microseconds; this is the time during which the analog gate G is open and charge accumulates on capacitor C due to the energy source V. The second time interval, τ, is

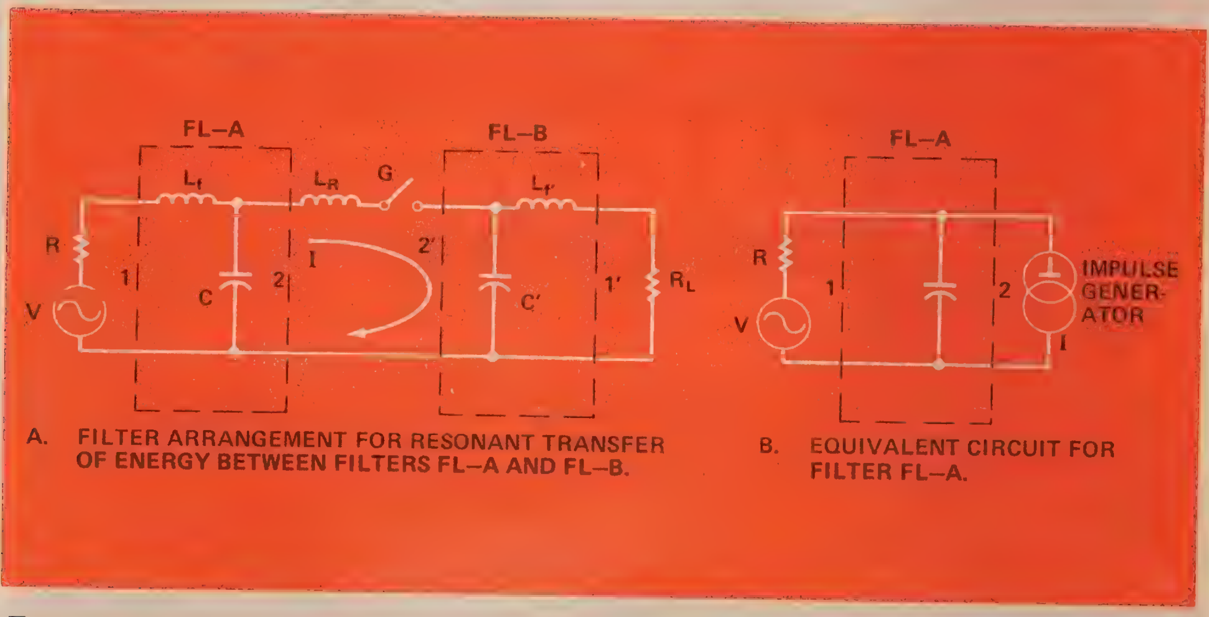


Figure 1. Circuit arrangement for resonant transfer operation.

approximately one microsecond. This is the period during which the analog gate G is closed. The inductor L_R is selected to resonate with capacitors C and C′ so that exactly one-half sine wave of current flow occurs during the one-microsecond closure of gate G. The series inductors assumed to be in the filters appear as open-circuits to the abrupt change in current produced during τ. Thus, during τ, the circuit effectively consists only of the series resonant circuit made up of C, C′, L_R, and G. Together, these elements can be considered the resonant transfer mechanism.

During the gate closure interval τ, the charge which existed on capacitor C is transferred to capacitor C′, thus reducing the charge on C to zero by the end of the closure interval. During the following T_s interval, while capacitor C is accumulating a new charge, capacitor C′ is being discharged through to the load resistor R. With proper filter design, the charge on C′ should be reduced to zero by the end of the T_s interval, when gate G closes.

Time-Domain Representation of Resonant Transfer

The timing sequence for closing analog gate G is shown in Figure 2A. As previously described, the gate is closed for τ microseconds out of each T_s microsecond interval. The signal

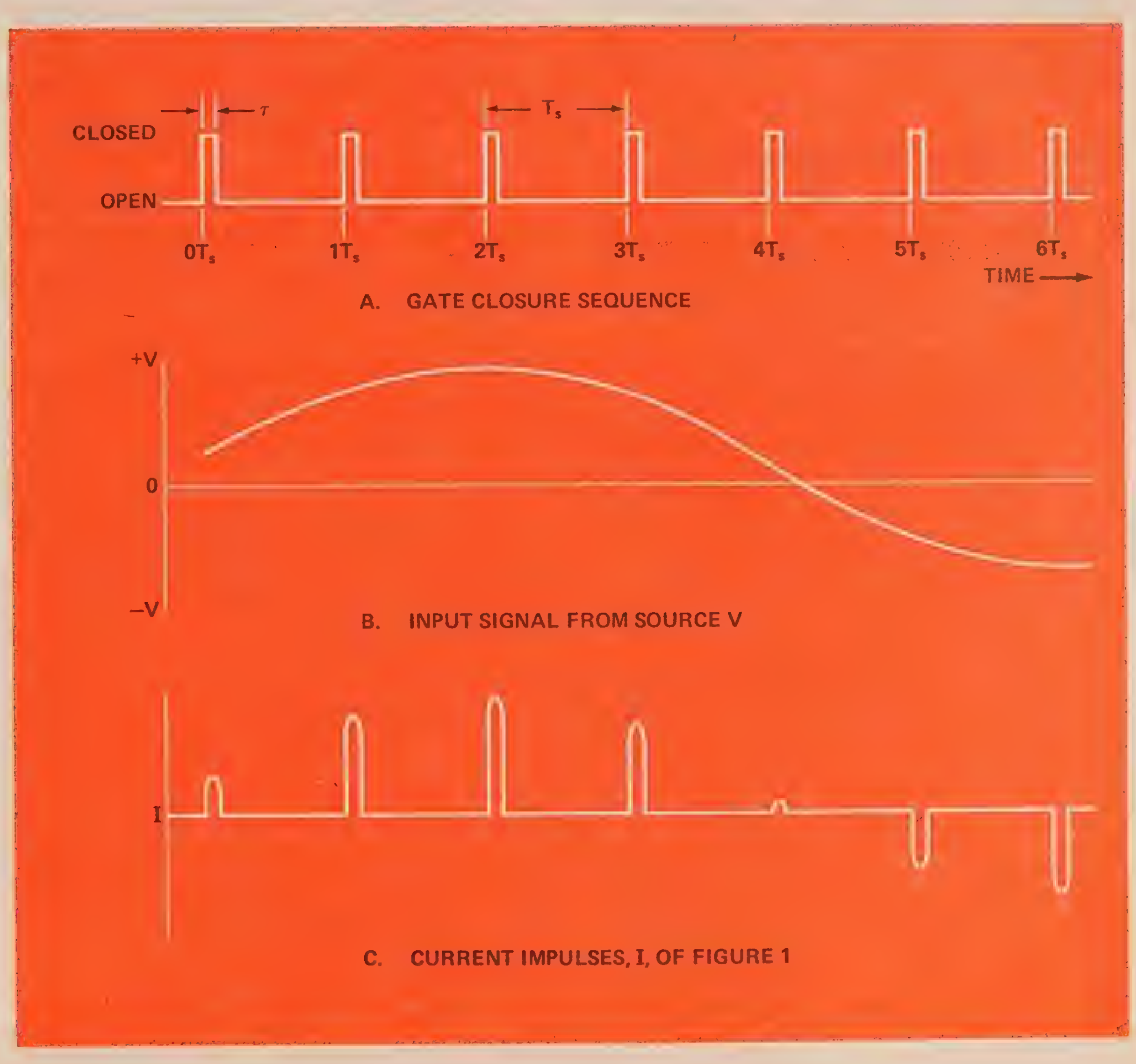


Figure 2. Timing for resonant transfer.

applied to the circuit by the source V is indicated in Figure 2B. At each switch closure interval the charge on capacitor C and, therefore, the magnitude of the current impulse (I) will be proportional to the signal amplitude. The resulting sequence of current impulses is shown in Figure 2C.

Because of the current impulse sequence appearing at the gated port (2′) of the filter, an equivalent representation of the resonant transfer process as applied to a single filter can be conceived. This is shown in Figure 1B. Here the source at port 2 is not a conventional generator; it is, rather, a special current source which produces the sequence of impulses shown in Figure 2C. This sequence of current impulses, in conjunction with the filters, has certain characteristics which are of interest.

Filter Impulse Response

As each current impulse is applied to the filter, it produces a voltage at the gated port which decays in an oscillatory manner. The general shape of the voltage waveform is that of a sinX/X pulse. While in actual operation the consecutive impulses at the gated port would be continuous, the voltage produced by three of these impulses would appear as shown in Figure 3. If the filter frequency response is designed properly, the response to each impulse will be such that the voltage wave will pass through zero at multiples of T_s subsequent to the impulse. If this condition holds, and the various multiples of T_s are examined, it will be noted that at any instant, t = kTs, only one impulse contributes to the magnitude of the voltage — the waveforms due to all other impulses are zero. In mathematical terms, the waveforms are said to be orthogonal at t = kTs. This type of resonant transfer filter operation might be termed "impulse response zero" operation.

One filter that meets the "impulse response zero" criterion which requires that the voltage response to a

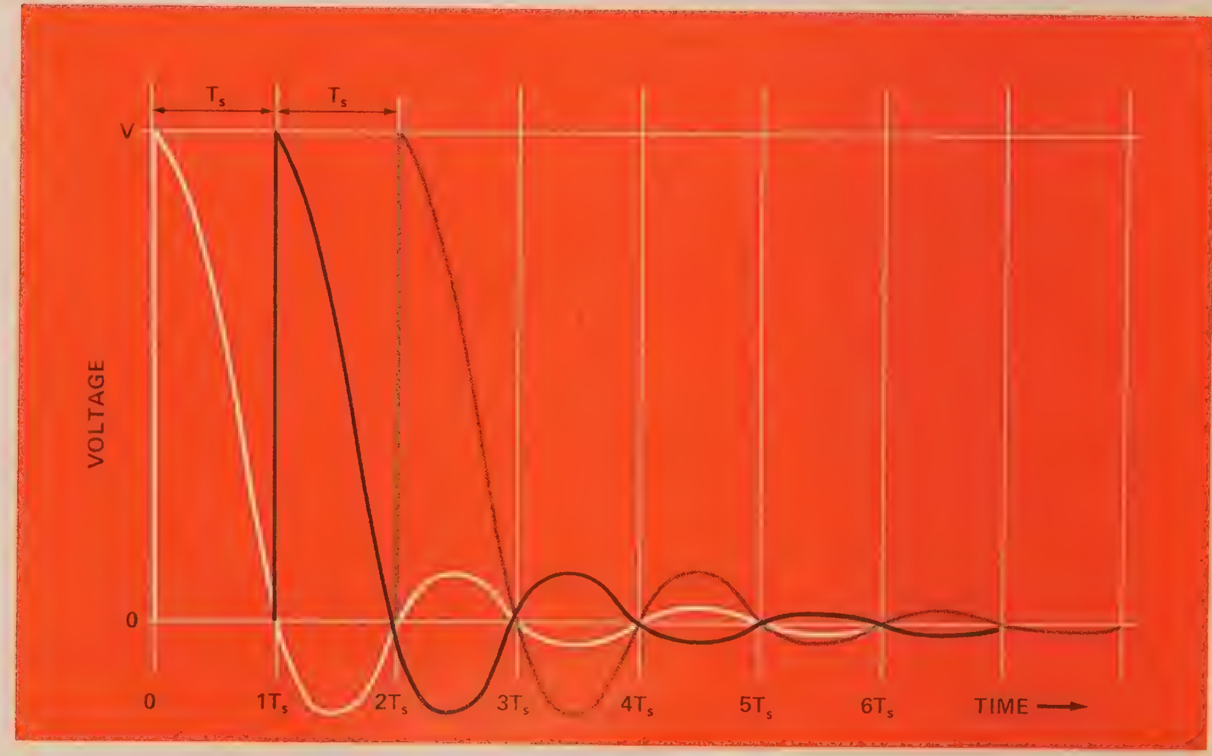


Figure 3. Voltage at gated port of filter due to three current impulses.

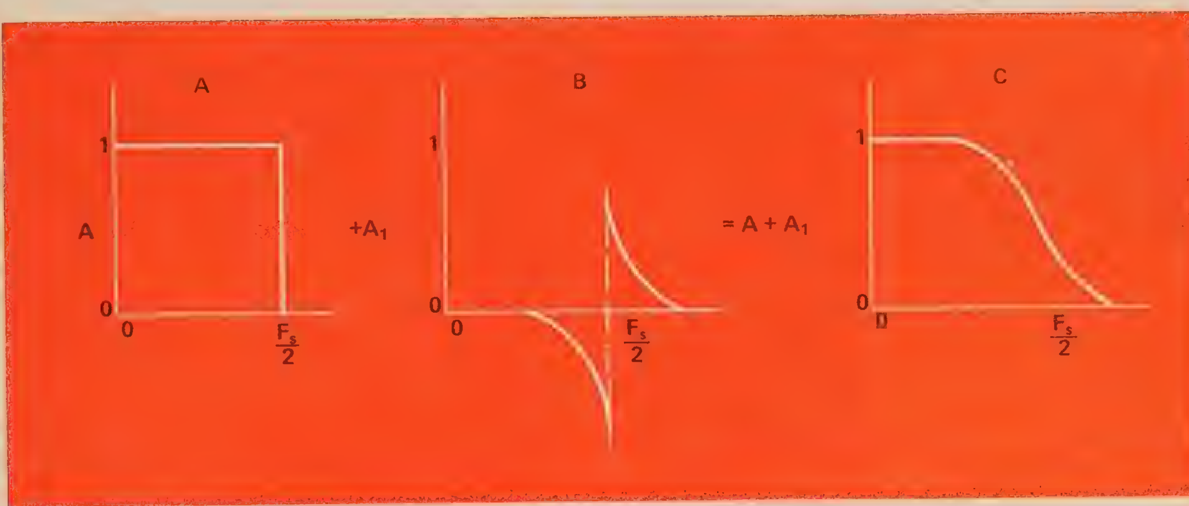


Figure 4. A characteristic having skew symmetry about $F_s/2$ as shown in (B) can be added to an ideal filter (A) characteristic to obtain an equal area curve having no abrupt transitions (C).

current impulse pass through zero at multiples of T_s is the "ideal" filter whose characteristic curve is shown in Figure 4A. This filter has a passband extending from 0 to $F_s/2$ Hz, where its stopband abruptly begins.

Unfortunately, because of the abrupt transition required from passband to stopband, such an ideal filter is physically unrealizable. However, a theorem by Nyquist, based on mathematical constructs, states that if a transmittance having the characteristic of skew-symmetry (as shown in Figure 4B) is added to the transmittance of an ideal filter (Figure 4A), the resulting filter transmittance will have the same zero crossings, in the time domain, as the ideal filter. A suitable choice for the skew-symmetric shape leads to a final filter response having the gradual passband-stopband transition (in the frequency domain) shown in Figure 4C. One filter characteristic shape which is arrived at in this manner is the "raised-cosine" filter shown in Figure 5. The frequency response and the impulse response of the raised-cosine filter of Figure 6 are shown in Figures 7 and 8, respectively, where it can be seen that this filter is readily realizable.

It is reasonable to assume that if all the energy is removed from the filter

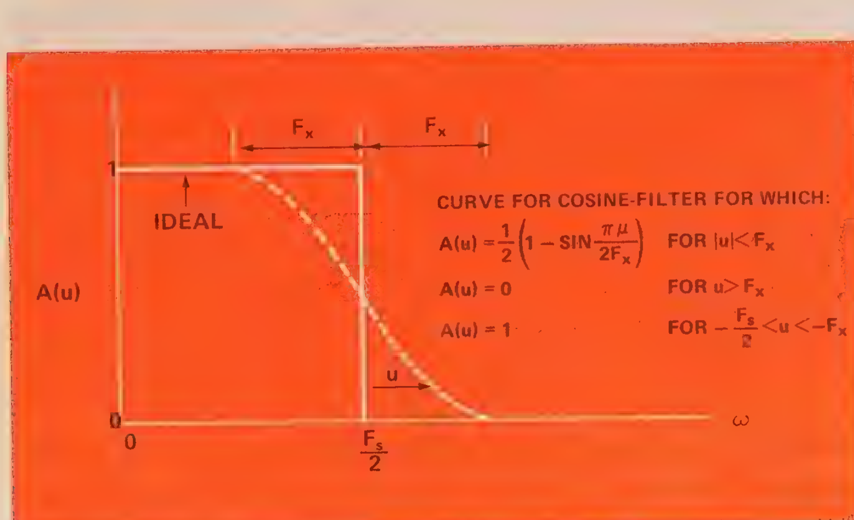


Figure 5. Comparison of characteristic curves of ideal low-pass filter and raised-cosine filter.

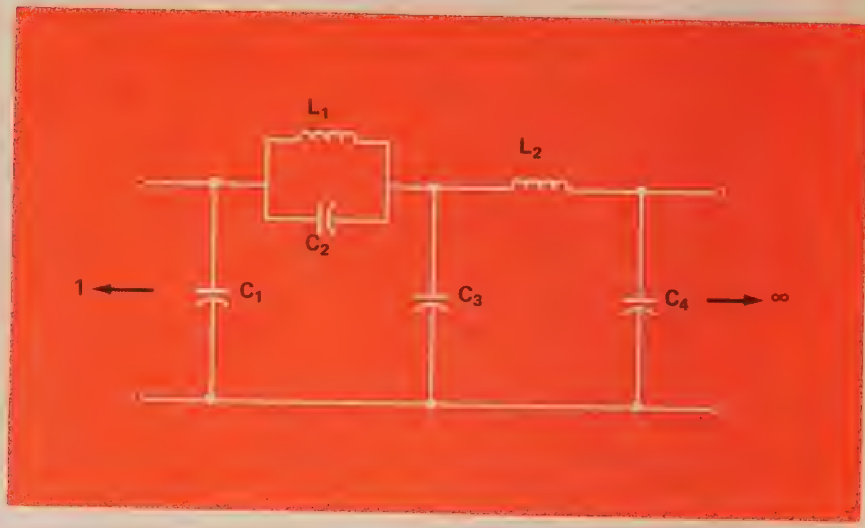


Figure 6. Schematic of a raised-cosine filter.

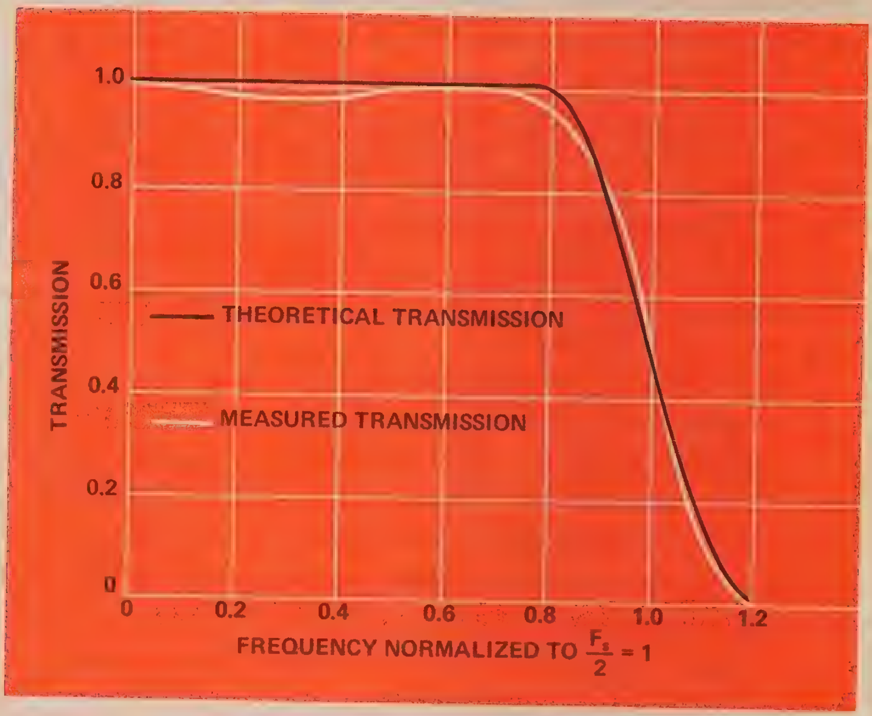


Figure 7. The frequency response of a raised-cosine filter with 25 dB attenuation at $F = 1.2F_s/2$.

at the time of an impulse, and that if none of this energy can be returned to the filter at subsequent switch closure intervals due to the zero-crossing criteria (subsequent switch closures will have zero voltage), then the resonant transfer technique is an inherently low-loss process. A loss on the order of 1.5 dB through a pair of filters connected as shown in Figure 1 is readily attainable.

Frequency Domain Response

In Figure 2C, the pulse sequence at the switched port of the filters is amplitude modulated by the input source V. This is presented pictorially in Figure 9, where A represents the frequency spectrum of the source and B shows the spectrum at the gated port. The input spectrum is repeated at the gated port as upper and lower sidebands spaced about multiples of the sampling frequency F_s. If the input frequency is designated as F_v, in Hz, the spectrum at the gated port is, then, $kF_s \pm F_v$ Hz, where $k = 0,1,2,3$, etc., represents the multiples of F_s. The point of interest is that filter FL-A of Figure 1 can be a

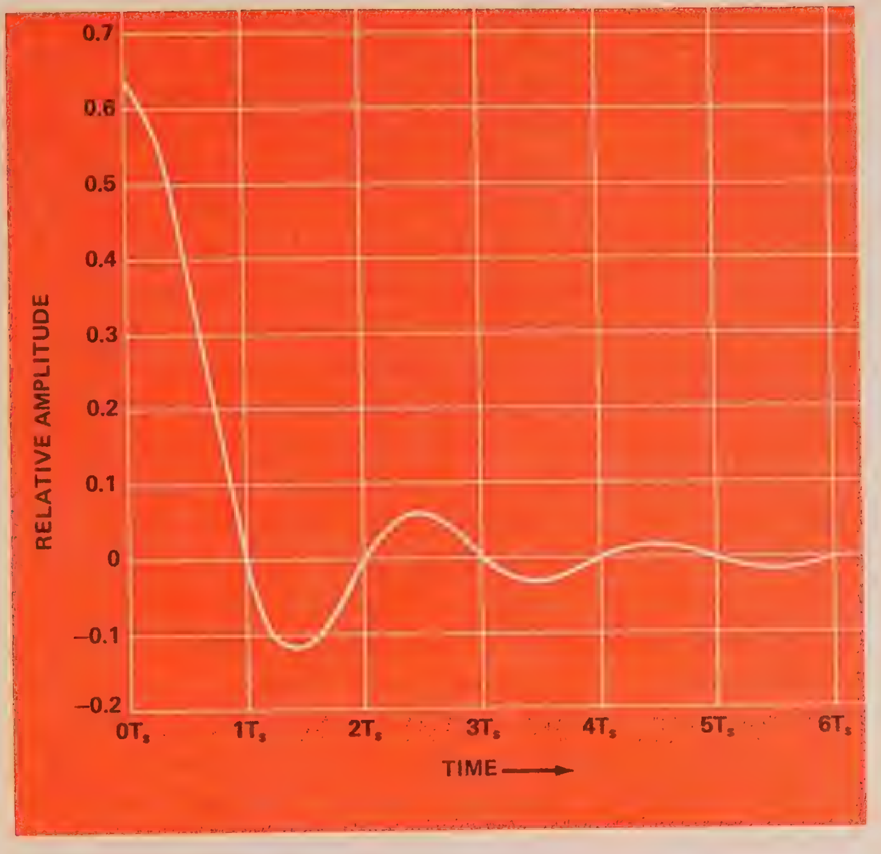


Figure 8. Impulse response of a raised-cosine filter.

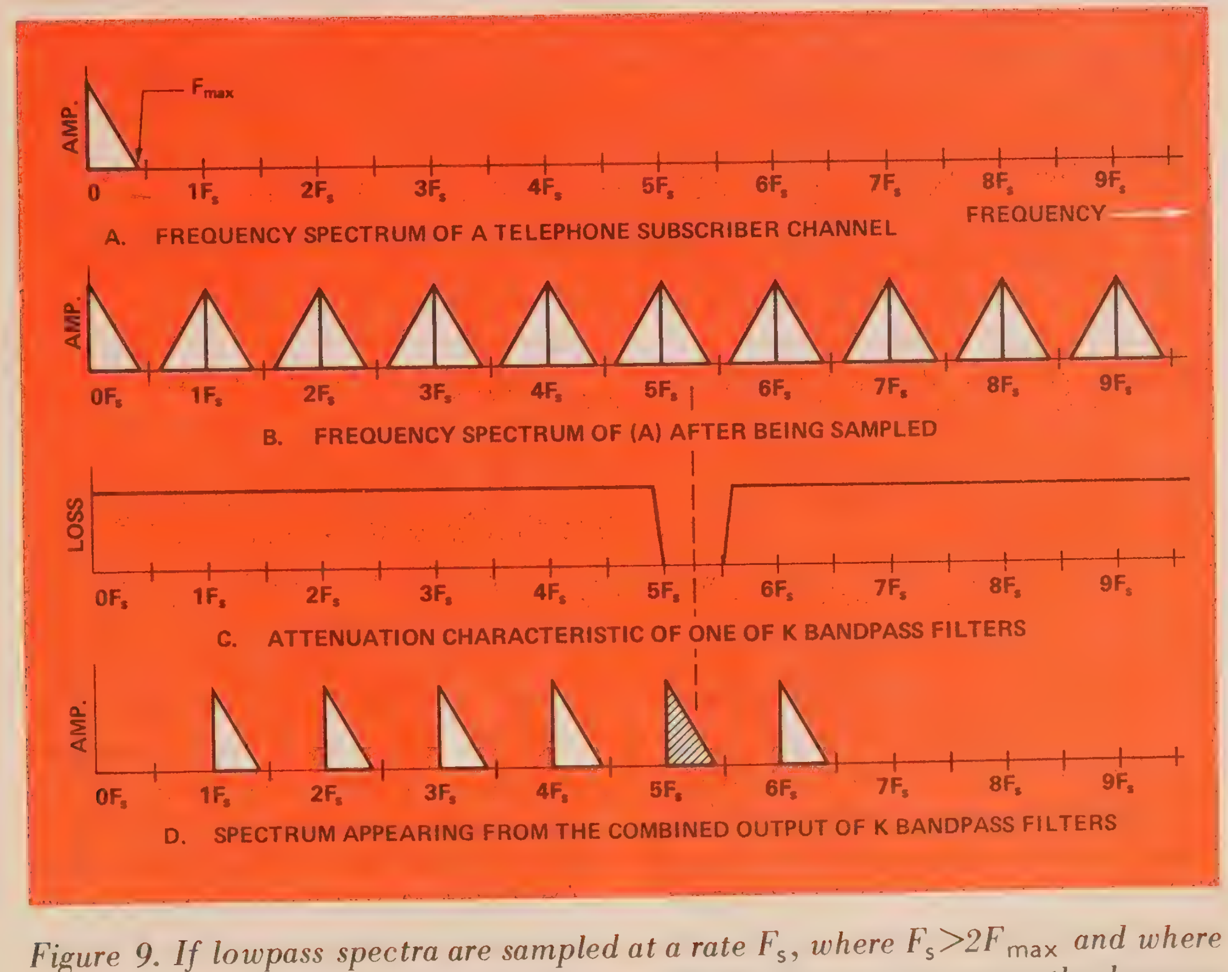


Figure 9. If lowpass spectra are sampled at a rate F_s, where $F_s > 2F_{max}$ and where F_{max} is the highest frequency of interest in the lowpass spectra, the lowpass spectra will be reproduced as sidebands about multiples of F_s.

lowpass filter while FL-B, as shown in Figure 9C can be a bandpass filter. The passband frequency of FL-B can be designed to pass a single sideband, as shown. This arrangement results in a zero-loss single sideband modulator.

If this process is repeated for several modulators, each having different passband frequency allocations, the several modulator outputs can be combined to form the mutlichannel, frequency-division multiplex line signal shown in Figure 9D. Two very significant differences appear in this modulation process compared with conventional single sideband modulators. Here, the modulator has a theoretical loss of 0 dB, and about 1.5 dB actual loss. A conventional modulator has a theoretical loss on the order of 3 dB and an actual loss on the order of 4.5 dB. Furthermore, a conventional modulator requires a carrier input frequency kF_s to generate each pair of sidebands ($kF_s \pm F_V$). Thus, six separate carrier frequencies would be required for conventional modulators, one for each modulator, to derive the spectrum of Figure 9D. With the resonant transfer technique, it is only necessary to apply the single pulse train shown in Figure 2A to all six modulators. This allows a substantial savings in system cost.

The Resonant Transfer 2/4-Wire Hybrid

The circuit of Figure 1 is, in addition to being a theoretically zero-loss configuration, bidirectional. That is, the source could just as well be placed at port 1′ and the load at port 1. In spite of the zero-loss characteristics of this circuit it may be used in applications where a high loss exists between an input or output port and a termination. It is then necessary to add gain while retaining the bidirectional transmission path. Since conventional amplifiers are unidirectional, this situation presents a problem. One possible solution is the use of "negative impedance" amplifiers, which are bidirectional but tend to oscillate unless the gain is quite low. The other possibility is to rearrange the circuit of Figure 1 to form a configuration which behaves just as a conventional 2/4-wire hybrid.

The resonant transfer 2/4-wire hybrid arrangement is shown in Figure 10. The filters may be all lowpass or a mixture of lowpass and bandpass as shown. In the operation of the circuit, a signal applied to the two-wire port charges a capacitor on the switched side of the lowpass filter. When gate G1 closes, the energy stored in this capacitor is discharged into the transmit bandpass filter using the resonant transfer process. Immediately after gate G1 opens, while the energy stored in the lowpass filter is still zero, gate G2 is closed. This causes transfer of energy from the receive bandpass filter to the lowpass filter. The interval which follows – from the opening of G2 until the closing of G1 – is very nearly equal to T_s. This choice of interval allows the energy applied to the lowpass filter from the receive bandpass filter to be transferred through the lowpass filter to the 2-wire termination. Thus, only the energy from the source at the 2-wire port is present at the switched side of the lowpass filter when gate G1 closes.

Generally, performance of the resonant transfer hybrid is very similar to conventional transformer hybrids. However, it offers several distinct advantages. The original objective in considering a hybrid was to be able to insert gain into the system. This is done as indicated in Figure 11, where a net gain in the West-East and in the East-West direction is desired. Insertion of the amplifiers as shown provides the required gain. Of course, it is

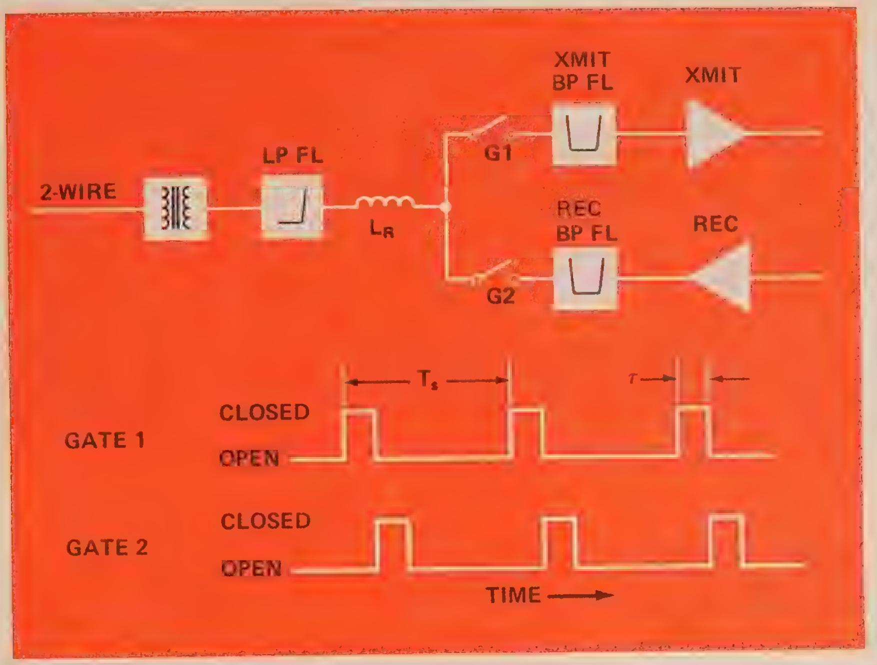


Figure 10. Resonant transfer 2/4 wire hybrid.

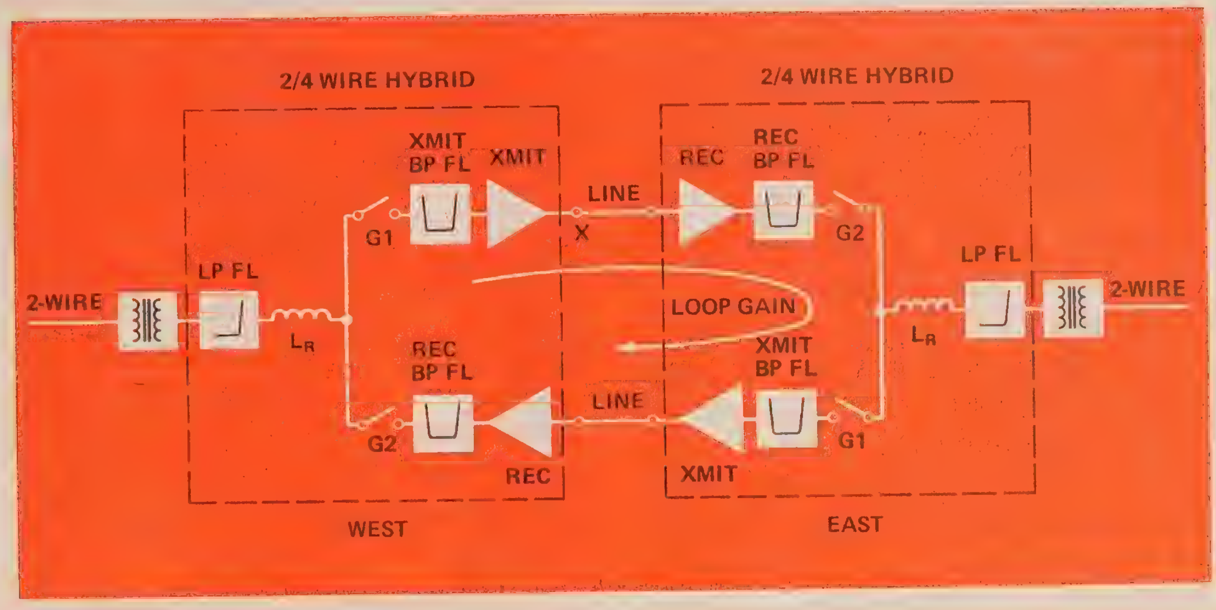


Figure 11. The use of 2/4 wire hybrids to allow insertion of amplifiers in transmission path.

necessary that the gain around the loop from any given point X back to X be less than 0 dB, or the circuit will oscillate; such oscillation is called "singing." It is the purpose of the hybrids to insert sufficient loss (called transhybrid loss) in the loop to allow this condition to be met and yet to insert very little loss in the direct transmission paths from West to East and East to West.

The resonant transfer hybrid has about 1.5 dB loss in the direct transmission paths; that is, from the hybrid 2-wire port to its transmit port and from receive port to two-wire port. The transhybrid loss – the loss from the receive port to the transmit port – is about 35-40 dB. By comparison, a conventional hybrid has a loss in the direct transmission path somewhere on the order of 4.5 dB and a realizable

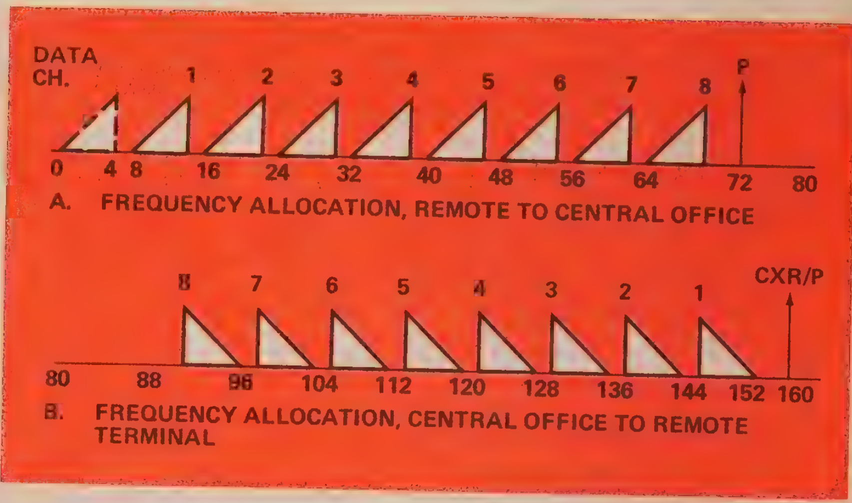


Figure 12. Line signals between central office and remote terminal.

transhybrid loss in the 35-40 dB range.

A conventional hybrid is a lattice or bridge type network which depends on balancing the bridge arms to achieve transhybrid loss. One arm of the bridge structure is a "balancing" network which must have an impedance characteristic that matches the impedance seen looking from the two-wire port into the two-wire line. This matching network is not required in the resonant transfer hybrid.

Advantages of the resonant transfer hybrid are: (a) it achieves the same transhybrid loss as a conventional hybrid while having 3 dB less loss in the transmission path, (b) it requires no balancing network, and (c) by using all lowpass filters it can be made to behave as a conventional vf-to-vf (voice frequency) hybrid, but by utilizing bandpass transmit and receive filters, it becomes both a hybrid and a modulator as well.

Resonant Transfer Frequency-Division Multiplex System

Using eight of the 2/4-wire hybrids with bandpass filters spaced at 8 kHz intervals in the frequency range from 8 to 64 kHz, the spectrum shown in Figure 12A can be formed. This basic 8-channel group allocation is used in the 48 channel multiplex system shown in Figure 13 for transmission in

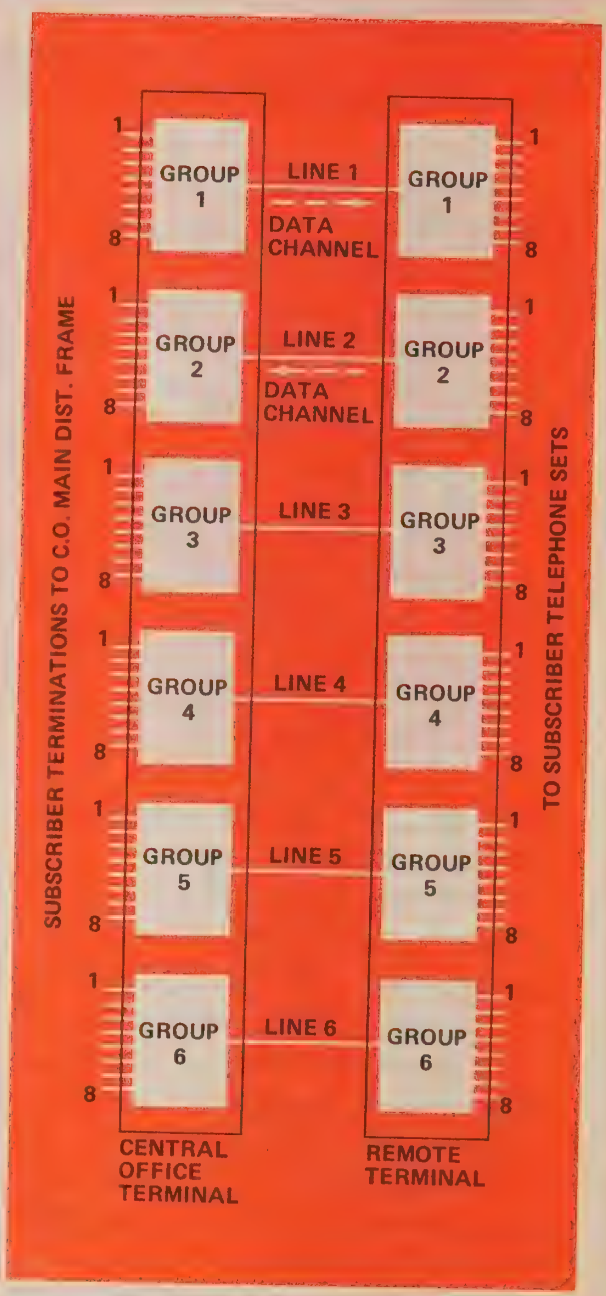


Figure 13. Arrangement of resonant transfer subscriber multiplex system.

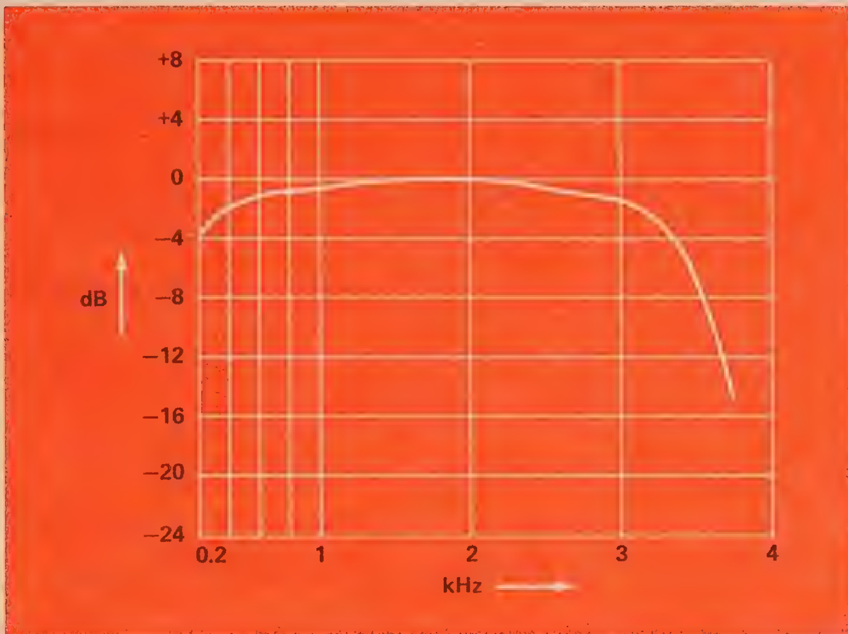


Figure 14. The channel response for a 64-68 kHz channel.

the remote terminal to central office terminal direction. The high frequency line signal shown in Figure 12B is used for transmission in the other direction; it is formed by modulating the spectrum of Figure 12A with 160 kHz and using the lower sideband. Both signals in Figure 12 are transmitted, in opposite directions, over a single physical cable pair between the two carrier terminals.

The system can be implemented in 8-channel increments to its ultimate 48-channel capacity, with one cable pair being required for each 8-channel group. Typical frequency response for the 64 – 68 kHz channel is shown in Figure 14.

The resonant transfer technique has possibilities in many aspects of communications, and, while its full potential has perhaps not as yet been realized, it holds the promise of affecting future trends in telecommunications.

BIBLIOGRAPHY

1. Bennett, W.R. and J. Davey. *Data Transmission.* New York: McGraw-Hill, 1965.

2. Getgen, L.E. "A 48-Channel Subscriber Multiplex System Using Resonant Transfer Modulation," *International Conference on Communications,* (June 1975).

3. Getgen, L.E. *The Use of Resonant Transfer for Combined Switching and Multiplexing.* Doctoral Thesis. Oregon State University, 1972.

4. "Hybrids," *The Lenkurt Demodulator,* January 1964.

INDEX TO VOLUME 3